TAMPEREEN TEKNILLINEN YLIOPISTO

JANNE VENHO

Verkkopohjaisen opiskelun lokitietoanalyysi

DIPLOMITYÖ

Tarkastaja: prof. Seppo Pohjolainen
Tarkastaja ja aihe hyväksytty
Luonnontieteiden tiedekuntaneuvoston
kokouksessa 06. helmikuuta 2013

TIIVISTELMÄ

TAMPEREEN TEKNILLINEN YLIOPISTO
Teknis-luonnontieteellinen koulutusohjelma
VENHO, JANNE: Verkkopohjaisen opiskelun lokitietoanalyysi
Diplomityö, 72 sivua, 19 liitesivua
Kesäkuu 2013
Pääaine: Matematiikka
Tarkastaja: professori Seppo Pohjolainen
Avainsanat: lokitietoanalyysi, klusterianalyysi, matematiikkajumppa, verkkopohjainen opiskelu

Tekniikan alan kasvava työllisyystarve on pakottanut yliopistot kasvattamaan opiskelijoiden sisäänottoa. Tämä on aiheuttanut suurempia keskeytysprosentteja, koska kaikkien opiskelijoiden lähtötaso ei ole enää yliopisto-opintojen vaatimalla tasolla varsinkaan matematiikan osalta. Ongelmaan kehitetään uusia tukitoimia, ja varsinkin internetpohjaiset opiskeluympäristöt ovat kehityksen kohteena. Yksi tällainen on Euroopan Komission vuonna 2009 käynnistämä Math-Bridge-projekti, jonka seurauksena saatiin käyttöön internetpohjainen Math-Bridge-ohjelmisto matematiikan itsenäistä kertaamista varten.

Tämä tutkimus on osa Tampereen teknillisen yliopiston Matematiikan laitoksen ja IISLabin opetuksen kehittämisen tutkimusta. Lähtökohtana tutkimukselle on Insinöörimatematiikka 1u -kurssin syksyllä 2011 aloittaneet opiskelijat. Opiskelijoille järjestettiin ennen kurssin alkua perustaitotesti, jonka perusteella heikoimmin menestyneet opiskelijat ohjattiin matematiikkajumppaan. Perustaitotestin tai matematiikkajumpan suorittaminen oli osa kurssin suoritusta.

Matematiikkajumppa toteutettiin Math-Bridge-ohjelmistolla, ja jumppaan osallistuneiden opiskelijoiden toiminnasta kerättiin lokitietoja jumpan aikana. Osaamista mitattiin perustaitotestin, jumpan jälkeisen jälkitestin sekä Insinöörimatematiikka 1u:n tentin avulla. Lisäksi opiskelijoista saatiin luokittelevaa tietoa opiskelijaprofiilikyselystä. Näiden tietojen avulla tutkittiin jumpan suorittaneiden opiskelijoiden toimintaa jumpan aikana, ja tutkittiin toiminnan yhteyttä testituloksiin.

Tutkimusmenetelminä käytettiin aluksi tilastotieteen perustyökaluja, joilla aineistoa kuvattiin ja analysoitiin. Tämän lisäksi opiskelijoita ryhmiteltiin klusterianalyysin avulla, jotta voitiin tunnistaa lokitietojen perusteella erilaisia työskentelytapoja matematiikkajumpassa.

Tutkimuksen perusteella löydettiin ryhmiä, joiden opiskelu matematiikkajumpassa oli päämäärätöntä eikä oppimista tapahtunut. Nämä opiskelijat ovat siis lokitietoanalyysin avulla havaittavissa, jolloin heille voitaisiin tarjota tarvittavia tukitoimia. Lisäksi havaittiin ne opiskelijat, joille matematiikkajumpasta sinällään oli hyötyä asioiden kertaamisessa.

ABSTRACT

TAMPERE UNIVERSITY OF TECHNOLOGY
Master's Degree Programme in Science and Engineering
VENHO, JANNE: Log data analysis of web-based learning
Master of Science Thesis, 72 pages, 19 Appendix pages
June 2013
Major: Mathematics
Examiner: Professor Seppo Pohjolainen
Keywords: log data analysis, cluster analysis, Mathematical Remedial Instruction, web-based learning

The need of new employment in the field of technology has forced universities to increase the intake of students. This has caused increasing drop-out rates because the starting levels of engineering students are often not on the required level for university studies. New measures of support are developed to work out the problem and especially web-based learning environments have been developed. One example of this kind of environment is the European Comissions Math-Bridge project. It started in 2009 and produced the Math-Bridge software for self-access learning.

This research is a part of research concerning improvement in the teaching of mathematics which is being researched in the Institute of mathematics and IISLab at Tampere University of Technology (TUT). The starting point for the study are the students that started the course Engineering mathematics 1u in the autumn of 2011. At the beginning of the course a Basic Skills Test (BST) was organized for the students. Those students who did not pass the BST had to participate in the Mathematics Remedial Instruction (MRI). Passing the BST or MRI was a compulsory part of Engineering mathematics 1u course.

Mathematics Remedial Instruction was carried out with Math-Bridge. Log data was collected during the MRI. In addition to log data, test results were collected from the BST before and after the MRI. Together with the exam results, the students' skills were evaluated. Additionally, classifying information like a learner profile was also obtained from the students. Using this information, the activity of students was studied and compared with test results.

Basic statistical tools were used as research methods at the beginning when data was described and analysed. In addition, students were grouped with cluster analysis to recognize different working methods in the MRI from log data.

Some useful groups were found on the strength of the study. There were students who studied aimlessly and did not learn mathematics effectively. With the help of log data analysis, suitable support can be offered to these students. In addition, students who really made use of the MRI and revised different fields of mathematics were also found from the grouping.

ALKUSANAT

Tämä diplomityö on jatkoa Tampereen teknillisen yliopiston Matematiikan laitoksen ja IISLabin opetuksen kehittämisen tutkimukselle. Työtä tehdessäni sain yhdistää taitojani matematiikan, kasvatustieteen ja ohjelmoinnin osalta. Sain myös paljon hyödyllistä oppia kielelliseen ilmaisuun sekä kuvien luomiseen.

Haluan kiittää professori Seppo Pohjolaista työni ohjaamisesta ja neuvoista tutkimuksen aikana. Lisäksi tahdon kiittää Jussi Kangasta, Ossi Nykästä sekä Teemu Mäkelää, jotka toimivat Math-Bridge-projektin työryhmässä kanssani lähes vuoden ajan. Heiltä sain paljon hyviä neuvoja ja kannustusta työtä tehdessäni. Kiitos myös koko Matematiikan laitoksen väelle loistavasta työilmapiiristä ja kaikesta avusta, jota olen työni aikana tarvinnut.

Kiitos työtovereilleni Ainolle, Jukka-Pekalle ja Tuomakselle mahtavista keskusteluista, väittelyistä sekä loistavasta työyhteisöstä. Erityiskiitos Juho Virpirannalle kieliasun tarkistuksesta sekä valtavasta määrästä apua ja neuvoa.

Tähän diplomityöhön kulminoituvat kaikki ne seitsemän vuotta, jotka olen Tampereen teknillisellä yliopistolla opiskellut. Matka on ollut pitkä, enkä olisi siitä selvinnyt ilman hyviä ystäviäni Jukka-Pekka Alankoa, Elina Miettistä ja Tuukka Peltolaa. Kiitos menneistä ja tulevista vuosista. Kiitos myös muille ystävilleni. Erityiskiitos tulevalle vaimolleni Leenalle kaikesta tuesta ja kannustuksesta opiskelujeni ja elämäni suhteen. Suurin kiitos tästä ja kaikesta kuuluu kuitenkin vanhemmilleni ja sisarelleni. Kiitos.

Tampereella 16. toukokuuta 2013

Janne Venho

SISÄLLYS

LYHENTEET JA MERKINNÄT

α, β	Regressioparametri
d	Absoluuttinen hajonta
$d(x, y)$	Havaintoarvojen välinen etäisyys
d_{UV}	Klustereiden U ja V välinen etäisyys
D	Kolmogorvin–Smirnovin testisuure
$\mathbf{D}$	Etäisyysmatriisi
e_i	Sovitetun regressiomallin jäännöstermi eli residuaali
f	Frekvenssi, esiintymistiheys
$f\%$	Prosenttiosuus
$f(X)$	Tiheysfunktio
F_0	Normaalijakauman kumulatiivinen suhteellinen frekvenssijakauma
Γ	Gammafunktio
H_0	Nollahypoteesi
H_1	Nollahypoteesin vastahypoteesi
K	Klusterikeskusten lukumäärä
μ	Satunnaismuuttujan keskiarvo
N	Otoskoko
ν	Vapausaste
p	Riskitaso
$q(f)$	Kvantiili prosenttipisteessä f
r_{ij}	Pearsonin korrelaatiokerroin muuttujille x_i ja x_j
r_{ij}^2 / R^2	Selitysaste
ρ_{ij}	Spearmanin järjestyskorrelaatiokerroin
s	Otoskeskihajonta
s_{ij}	Kovarianssi
s^2	Otosvarianssi
S_N	Otoksen kumulatiivinen suhteellinen frekvenssijakauma
σ	Satunnaismuuttujan keskihajonta
σ^2	Satunnaismuuttujan varianssi
t	t-jakauman testisuure
x, y	Muuttuja
x_i, y_i	Muuttujan havaintoarvo

$\bar{x}$	Otoskeskiarvo
X	Satunnaismuuttuja
$\hat{y}_i$	Sovitettu lineaarinen malli
z	Standardoitu arvo
CPT	Conditional Probability Table; ehdollisen todennäköisyyden taulukko
IISLab	Intelligent Information Systems Laboratory
IMa1u	Insinöörimatematiikka 1u -kurssi
ITS	Intelligent Tutoring System; älykäs opetusjärjestelmä
LaMa1u	Laaja matematiikka 1u -kurssi
MIT	Massachusetts Institute of Technology
Python	Ohjelmointikieli
SSE	Sum of squares of errors; virheen neliöiden summa
SST	Total sum of squares; todellinen neliöiden summa
TTY	Tampereen teknillinen yliopisto

1. JOHDANTO

Nykyaikainen eurooppalainen talouselämä nojautuu vahvasti korkealaatuiseen työvoimaan varsinkin tekniikan alalla. Kasvava tarve uusille insinööreille on pakottanut yliopistot kasvattamaan opiskelijoiden sisäänottoa, joka on kasvattanut myös opiskelijoiden opintojen keskeytysprosentteja. Tämä johtuu opiskelijoiden heikosta lähtötasosta sekä huonosta valmistautumisesta opiskelun vaatimuksiin. [5]

Tampereen teknillisellä yliopistolla (TTY) on panostettu opiskelijoiden keskeyttämisen ehkäisemiseen monilla tavoin. Matematiikan osalta tukitoimena tarjotaan uusille opiskelijoille opintojen alussa matematiikan perustaitotesti, jolla pyritään kartoittamaan opiskelijoiden matematiikan opiskelun lähtötasoa ja testin perusteella tarjoamaan tukea matematiikan opiskelussa sitä tarvitseville. Tätä tarkoitusta varten on kehitetty matematiikkajumppa, joka toimii tämän tutkielman tutkimuskohteena. TTY:n Matematiikan laitoksen ja IISLabin (Intelligent Information Systems Laboratory) aikaisemmissa tutkimuksissa perustaitotestistä ja matematiikkajumpasta on aikaisemmin tutkittu opiskelijoiden asenteita, motivaatiota sekä orientaatioita matematiikan opiskelussa.

Tutkimuksessa käytetään syksyn 2011 matematiikkajumpan aikana kerättyjä lokitietoja opiskelijoiden toiminnasta Math-Bridge-ohjelmistossa, joka toimi alustana matematiikkajumpalle. Tietokoneavusteinen ja internetpohjainen opiskelu on koettu hyödylliseksi apuvälineeksi asioiden kertaamisessa ja rutiinien harjoittelussa. Lokitietoja voidaan yleisesti hyödyntää monella eri tavalla sekä ohjelmistojen kehittämiseen että opiskelijoiden tutkimiseen. Tässä tutkielmassa hyödynnetään lokitietoja ensisijaisesti opiskelijoiden toiminnan tutkimiseen sekä jumppatehtävien kehittämiseen.

Tutkimuksen tavoitteena on löytää ne opiskelijat, joille on hyötyä matematiikkajumpasta. Toisaalta halutaan tutkia voidaanko lokitietojen avulla havaita sellaiset opiskelijat, jotka eivät hyödy matematiikkajumpasta vaan tarvitsevat muitakin tukitoimenpiteitä.

2. LÄHTÖKOHDAT JA TEORIA

Tietokoneiden ja varsinkin internetyhteyksien yleistyessä kotitalouksissa ja kouluissa 1990-luvulla alkoivat myös verkkopohjaiset oppimisympäristöt osoittautua tärkeäksi kehittämisen kohteeksi. Kasvatustieteilijät ovat historian saatossa reagoineet nopeasti uusien teknologioiden kehittymiseen haaveillen niiden hyödyntämisestä opetuksessa ja oppimisessa. Tämä on kuitenkin usein hiipunut alun innostuksen jälkeen, eikä uusista teknologioista ole tullut merkittävää lisää opettamiseen ja oppimiseen. Opettajien ensisijainen tehtävä on pohtia, kuinka oppilaat oppisivat mahdollisimman hyvin opiskeltavan asian ja vasta tämän jälkeen punnita erilaisia vaihtoehtoja asian esittämiseen. Uusia teknologioita ei oteta käyttöön, elleivät ne tarjoa oppimisen kannalta jotakin uutta ja parempaa. [1]

Tämän luvun tarkoituksena on esitellä erilaisten internet- eli verkkopohjaisten järjestelmien tutkimusta ja kehittämistä sekä Tampereen teknillisen yliopiston tarjomaa matematiikan tukitoimintaa yliopistossa aloittaville opiskelijoille. Verkkopohjaista opiskelua on pyritty hyödyntämään nimenomaan opiskelun tukena luomalla erilaisia kertaus- ja tukimateriaaleja verkkoon. Verkkopohjaisiin opetusympäristöihin on lisätty kattavan itseopiskelumateriaalin sekä tehtävien lisäksi myös erilaisia oppaita tai tuutoreita, jotka seuraavat automaattisesti käyttäjän toimia järjestelmässä ja tarjoavat tälle ongelmatilanteissa apua ja taitojen kehittyessä esimerkiksi vaikeampia tehtäviä. Yksi näistä järjestelmistä on tutkimuksessa käytetty Math-Bridge.

2.1 Math-Bridge

Euroopan Komission Information Societyn eContentplus-ohjelma käynnisti vuonna 2009 Math-Bridge-projektin, jonka tavoitteena on toimia eräänlaisena siltakurssina toisen ja kolmannen asteen oppilaitosten välillä mahdollistaen vaadittavien matematiikan taitojen yksilöllisen kertaamisen ja opiskelun yliopisto-opintojen alussa. Projekti päättyi vuoden 2012 tammikuussa ja sen lopputuloksena saatiin verkkopohjainen oppimisalusta Math-Bridge. [5]

Math-Bridge-projektin tavoitteena oli kerätä opiskelumateriaalia Euroopan laajuisesti ja koostaa ne yhtenäisiksi matematiikan rakenteiden mukaisiksi oppimisobjekteiksi. Tällä tavoin mahdollistettiin yksilöllisempi opetus kuin tavanomaisissa tukikursseissa. Math-Bridge-projekti käytti oppimisalustan pohjana älykästä opetus-järjestelmää (ITS – Intelligent Tutoring System) nimeltä ActiveMath, joka ottaa huomioon opiskelijan kehittymisen, jolloin se voi tarjota räätälöityjä tukitoimia esimerkiksi kasvattaen tehtävien vaikeustasoa käyttäjän oppimistason perusteella. [3]

Oppimisobjekteista, kuten interaktiivisista tehtävistä ja niihin liittyvästä teoriaosuudesta, voidaan koostaa erilaisia tukikursseja ja -kirjoja. Kaikki materiaalit ja tehtävät ovat järjestelmässä seitsemällä eri kielellä: englanniksi, saksaksi, suomeksi, hollanniksi, ranskaksi, espanjaksi sekä unkariksi. Tämä mahdollistaa sekä Euroopan laajuisen käytön että matematiikan opiskelun vieraalla kielellä.

2.2 Perustaitotesti ja matematiikkajumppa

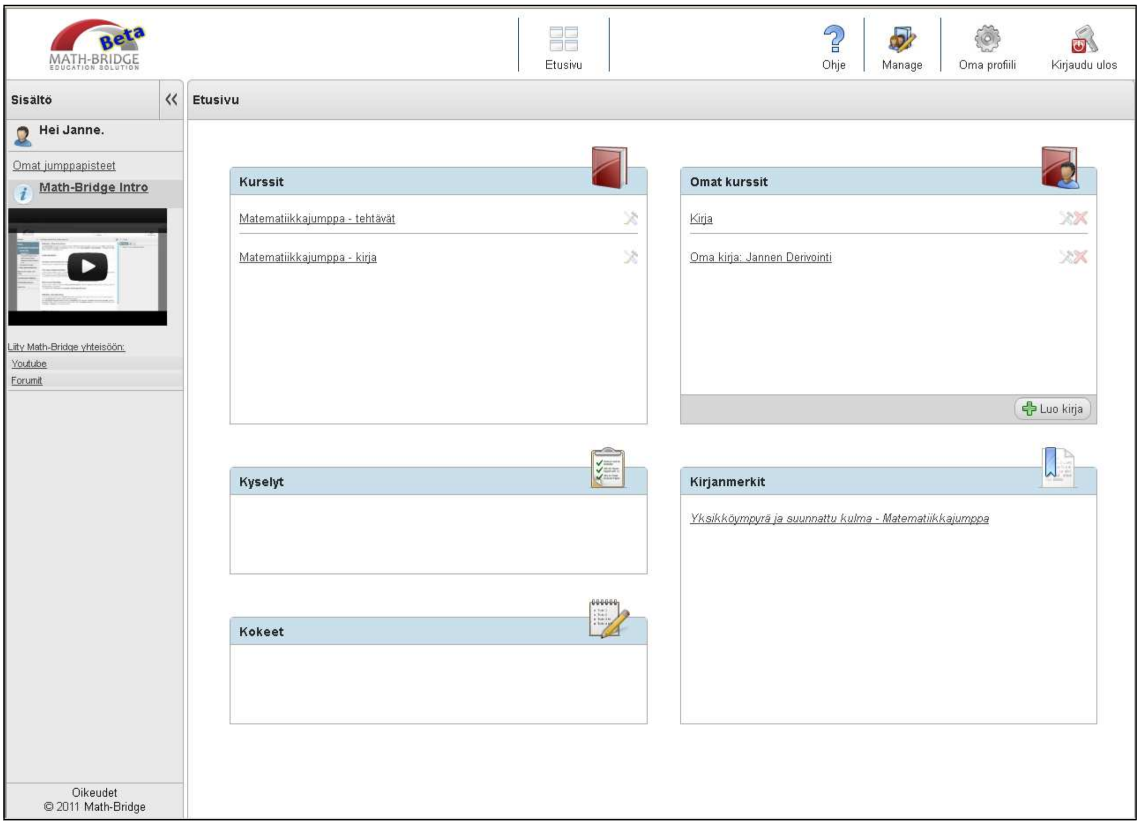
Matematiikan perustaitotesti on järjestetty Tampereen teknillisellä yliopistolla vuodesta 2002 saakka. Testi on tarkoitettu ensimmäisen vuosikurssin opiskeilijoille testaamaan lukiomatematiikan aihealueiden osaamista. Käytännössä perustaitotesti on pakollinen osa matematiikan perusopintoihin kuuluvia kursseja Insinöörimatematiikka 1u (myöhemmin IMa1u) sekä Laaja matematiikka 1u (LaMa1u). Aluksi testi järjestettiin ainoastaan kurssin IMa1u suorittajille, mutta myöhemmin myös kurssin LaMa1u suorittajat on otettu testin piiriin. Testissä on 16 lukiomatematiikan tehtävää, ja testi on suoritettu vuodesta 2011 alkaen Math-Bridge-ohjelmistolla.

Perustaitotestissä vähiten oikeita vastauksia antaneet opiskelijat ohjataan matematiikkajumppaan, jonka tarkoituksena on auttaa opiskelijaa kertaamaan vaadittavia matematiikan osa-alueita. Lisäksi ne kurssien IMa1u ja LaMa1u opiskelijat, jotka eivät osallistuneet perustaitotestiin voivat korvata puuttuvan testin suorittamalla matematiikkajumpan. Matematiikkajumppa (myöhemmin myös jumppa) suoritetaan kuten perustaitotesti, Math-Bridge-ohjelmistolla. Math-Bridgeen on jumppaa varten tuotettu Matematiikkajumppa-kirja sekä koottu kirjasta löytyvät tehtävät omaksi tehtäväpaketikseen. Tehtäviä on 71 kappaletta, ja opiskelijan on saatava kaikkiin tehtäviin oikea ratkaisu suorittaakseen jumppa hyväksytysti. Tehtävät ovat lukiomatematiikan tasoisia ja aiheet on jaoteltu kymmeneen eri osa-alueeseen: derivointi, yhtälönratkaisu, lausekkeet, epäyhtälöt, integrointi, raja-arvot, logaritmit, luvut, potenssifunktiot sekä trigonometriset funktiot. Syksyn 2011 matematiikkajumpan toteutusta ja suorittamista esitellään tarkemmin luvussa 3.

2.2.1 Math-Bridge matematiikkajumpassa

Opiskelijat kirjautuvat Math-Bridgeen TTY:n intranet-tunnuksilla, jolloin opiskelijoiden tunnistaminen on helppoa ja suoritukset voidaan kirjata oikeille henkilöille. Kuvassa 2.1 on esitetty Math-Bridgen etusivunäkymä sisäänkirjautuneelle opiskelijalle.

Etusivunäkymän vasempaan palstaan on lisätty linkki “Omat jumppapisteet”,



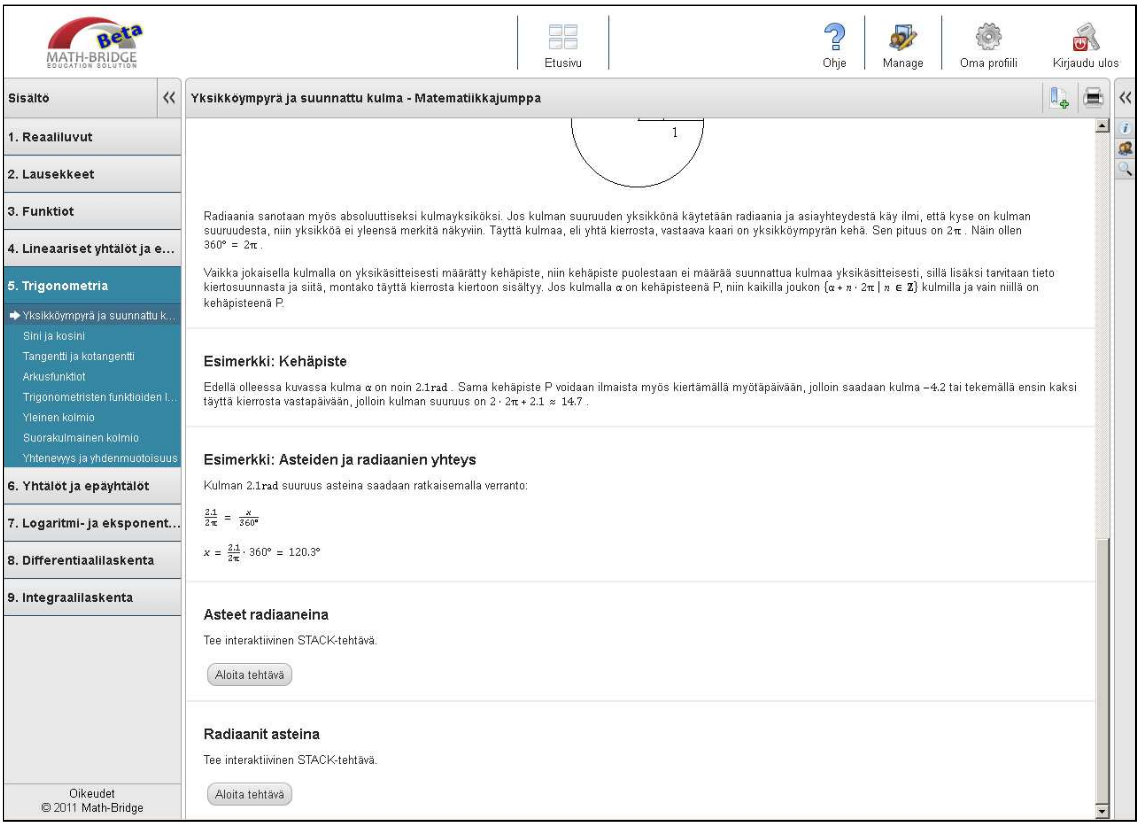
Kuva 2.1: Math-Bridgen etusivunäkymä kirjautuneelle käyttäjälle.

jonka avulla jumppaaja voi tarkastella suorittamiaan jumppatehtäviä. Tätä ominaisuutta ei ollut käytössä vielä syksyn 2011 matematiikkajumpassa, minkä vuoksi opiskelijoille ilmoitettiin suoritetuista tehtävistä jumpan verkkosivujen kautta määrävälein päivitetyssä pdf-dokumentissa.

Kuvassa keskellä olevissa laatikoissa tärkeimpänä näkyy “Kurssit”-osio, missä on näkyvissä matematiikkajumpan tehtävät sekä kirja. Näiden avulla opiskelija pääsee suorittamaan tehtäviä sekä opiskelemaan tarvittavaa teoriaa. Lisäksi opiskelija on voinut lisätä joitain eri kirjojen kohtia kirjainmerkeikseen, jolloin esimerkiksi opiskelun jatkaminen aiemmin keskeytyneestä kohdasta on helpompaa. Näiden lisäksi kursseihin saattaa sisältyä kyselyitä ja kokeita. Opettajan tunnuksilla voidaan koota myös omia kirjoja ja kursseja. Näitä tunnuksia ei tutkimuksen kohderyhmän opiskelijoilla ole, mutta tämä ominaisuus Math-Bridgestä mainittakoon.

Math-Bridgessa luotu kurssi koostuu erilaisista oppimisobjekteista, joita ovat esimerkiksi johdannot, motivoinnit, määritelmät, esimerkit ja interaktiiviset tehtävät. Kuvassa 2.2 on esitetty kuvakaappaus matematiikkajumpan kirjan sivulta. Kuvassa nähdään vasemmalla navigointipalsta, jonka avulla voidaan selata kirjan eri osioita helposti.

Avatessaan tehtävän Math-Bridge generoi kyseiselle tehtävälle uudet parametrit, jolloin tehtävä on hieman erilainen eri käyttökerroilla. Kuvassa 2.3 on esitetty näky-



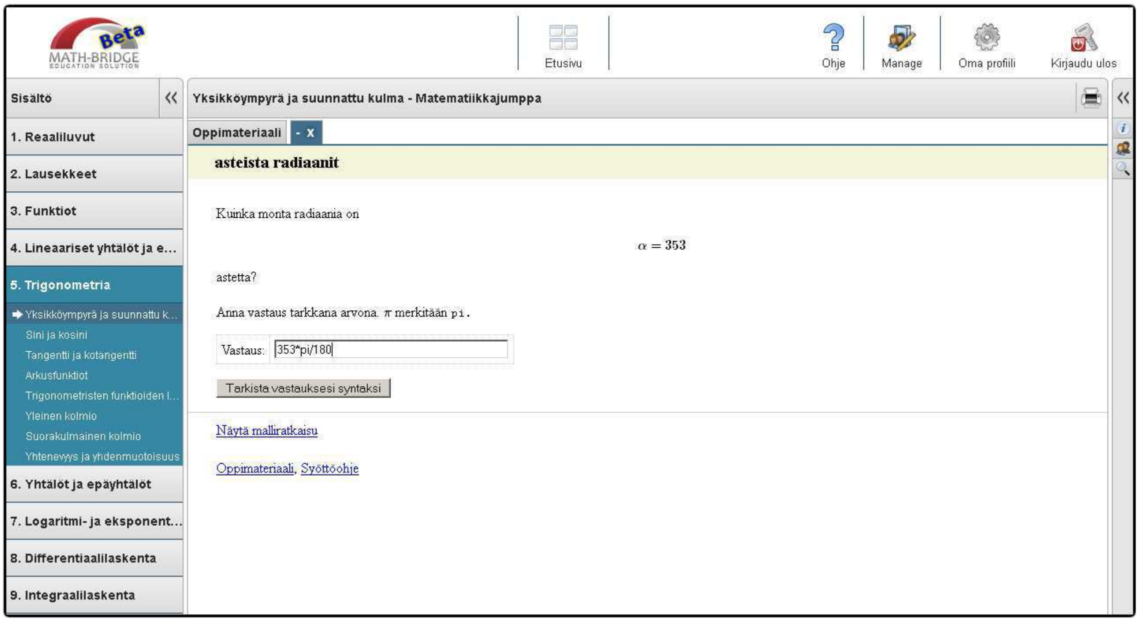
Kuva 2.2: Math-Bridgen kirjanäkymä.

mä tehtävän avaamisen jälkeen. Tehtävänannon jälkeen on "Vastaus"-kenttä, johon opiskelija syöttää saamansa vastauksen. Tämän jälkeen opiskelija voi tarkistaa vastauksensa syntaksin, jonka jälkeen vastausta voi vielä muuttaa, jos opiskelija ei ole tyytyväinen syöttämäänsä syntaksiin. Lisäksi ongelmatilanteissa opiskelija voi katsoa malliratkaisua, opiskella lisää oppimateriaalista tai katsoa syöttöohjetta syntaksia varten. Tarkemmin erilaisia näkymiä tehtävänratkaisun eri vaiheista on esitetty liitteessä A.

2.3 Lokitiedot

Verkkopohjaisissa järjestelmissä ohjelmisto on verkon yli käyttöä varten asennettu jollekin internetpalvelimelle. Käyttäjän käyttäessä järjestelmää esimerkiksi internetselaimen avulla selain lähettää palvelimelle pyyntöjä, kuten uuden sivun tai tehtävän lataaminen järjestelmästä. Palvelimet tallentavat automaattisesti lokitietoja näistä pyynnöistä ja lisäävät tietoihin aikaleiman, jolloin käyttäjien toimia voidaan seurata ja tutkia kronologisessa järjestyksessä. Lokitietoja käytetään opetuksen ja oppimisen tutkimuksessa sekä opiskelijoiden käytöksen tutkimiseen että järjestelmien ja varsinkin interaktiivisten tuutorien kehittämiseen ja testaamiseen.

Young-Jin Lee [10] esittelee tutkimuksessaan laskennallisen tavan, jonka avulla



Kuva 2.3: Math-Bridgen tehtävänäkymä.

voidaan arvioida opiskelijan taitoja ongelmanratkaisussa hyödyntäen verkkopohjaisen oppimisympäristön lokitietoja. Tutkimuksessa arvioitiin 407 korkeakouluopiskelijaa (college), jotka ratkaisivat 61 fysiikan tehtävää. Lokitiedot saatiin Massachusettsin teknillisen korkeakoulun (MIT – Massachusetts Institute of Technology) kehittämästä MasteringPhysics -tutorointijärjestelmästä, ja laskennallista menetelmää käytetään juuri tämän tuutorin kehittämisessä. Menetelmän avulla voidaan tutkimuksen mukaan tunnistaa oikea hetki antaa opiskelijalle hänen tarvitsemaansa tukea, jolloin voidaan parantaa merkittävästi oppimisen tehokkuutta. Vastaavanlaista tuutorointijärjestelmää tutki myös Kurt VanLehn et al. [16] ja he saivat hyviä tuloksia oppimisen tehostumisesta käytettäessä järjestelmää kotitehtävien tekemiseen.

Myös Arroyo ja Woolf [2] kehittivät uuden sukupolven tuutorointijärjestelmää, joka huomioi käyttäjien toiminnan järjestelmässä ja kehittää automaattisesti itseään tekemään pedagogisia päätöksiä käyttäjän tunteiden, järjen ja oppimistilanteen perusteella. He tutkivat 230 15–17-vuotiasta yläkoulun ja lukion opiskelijaa (high school). Opiskelijat tekivät esitestin, käyttivät kahdesta kolmeen tuntia Wyang Outpost -järjestelmää ja tekivät tämän jälkeen jälkitestin. Lopuksi he vastasivat kyselyyn, jolla mitattiin heidän piileviä asenteitaan sekä oppimista. Järjestelmän lokitiedoista saatiin muuttujia kuten virheet, aika ja avunpyynnöt. He havaitsivat, että opiskelijat, jotka pyysivät apua, oppivat todennäköisemmin ja ettei tehtävässä käytetty aika korreloinut positiivisesti oppimisen kanssa. Lisäksi he havaitsivat, että paljon virheitä järjestelmässä tehneiden opiskelijoiden parannus testien välillä oli pienempää ja suuri osa näistä vastasi haluavansa vain päästä eroon tehtävästä ilman perusteellisempaa syventymistä. He käyttivät tutkimuksessa ehdollisen todennäköisyyden taulukkoa (CPT – Conditional Probability Table), jossa aineisto jaetaan muuttujien mediaanien suhteen kahteen mahdollisimman yhtä suureen osaan. Jakoa toistetaan aina jaon tuloksena saaduille ryhmille eri muuttujilla ja tarkastellaan haluttujen jakojen jälkeen opiskelijoiden jakautumisen todennäköisyyttä näihin ryhmiin. Näistä havaittiin, että opiskelijat, jotka halusivat haasteellisia tehtäviä, käyttivät enemmän aikaa tehtävien tekemiseen kuin ne, jotka eivät haastetta kaivanneet. Samoin toimivat myös opiskelijat, jotka sanoivat pelkäävänsä vääriä vastauksia. Oppimisen kannalta CPT:n avulla havaittiin, että huonosti alkutestissä pärjänneille opiskelijoilla avun etsiminen oli tärkeä tekijä oppimisprosessissa, eivätkä hyvinkään alkutestissä menestyneet opiskelijat oppineet, jos he eivät käyttäneet riittävästi aikaa tehtäviin, joissa apua oli tarjolla.

Verkkopohjaisissa oppimisympäristöissä voidaan erilaisten tekstien, kuvien, animaatioiden sekä pelien avulla tuoda opittavaa asiaa mielekkäästi käyttäjän opiskeltavaksi. Tällä tavalla voidaan ottaa huomioon erilaisia oppijoita ja tarjota tukea heille tarvittaessa. Opiskelijoiden motivaatio on kuitenkin tekijä, joka vaikuttaa suuresti oppimistuloksiin. Cocea ja Webelzahl [4] tutkivat 20 käyttäjän lokitietoja tunnistaakseen opiskeluun sitoutuneet ja sitoutumattomat käyttäjät. He käyttivät HTML-Tutor-ympäristöä, jonka lokitiedoista he saivat esiin muun muassa käyttäjien tekemien tehtävien tilastoja sekä teoriasivujen lukemiseen käytetyn ajan. He käyttivät aikatietoja ja testituloksia muuttujina jakaessaan päätöspuun avulla käyttäjät sitoutumattomiin ja sitoutuneisiin. Vaikka pieni otos ei annakaan mahdollisuutta tulosten yleistämiseen, niin he havaitsivat ajankäytössä ja tehtävien osaamisessa tiettyjä raja-arvoja, joiden avulla voidaan päätellä opiskelijoiden sitoutumistasoa. Suurempi ajankäyttö koettiin paremmaksi sitoutumiseksi, mikä myös Arroyon ja Woolfin [2] mukaan lisäsi oppimista.

2.4 Opiskelijaprofiilit

TTY:n perustaitotestin alussa opiskelijoilta kysyttiin heidän opiskelutavoistaan antamalla vaihtoehdoiksi viisi erilaista opiskelijaprofiilia. Näistä opiskelijan tuli valita omaa matematiikan oppimista parhaiten kuvaava vaihtoehto. Profiilit ja kysymykset pohjautuvat Tampereen teknillisen yliopiston matematiikan laitoksen aikaisempaan tutkimukseen [15], jossa tutkittiin opiskelijoiden opiskelutapoja ja -motivaatioita. Alla on esitelty tutkimuksessa käytetyt profiilien nimet ja niitä vastaavat perustaitotestin yhteydessä annetut väittämät.

Pintasuuntautuneet oppijat: Kiinnostukseeni matematiikkaa kohtaan vaikuttaa enemmän koulutusohjelma kuin oma mielenkiinto. Lasken tehtävän usein samalla tavalla kuin se on esitetty kirjassa tai tunnilla, enkä yleensä mieti omaa menetelmää tehtävän ratkaisemiseksi. Kykenen oppimaan matematiikkaa kopioimalla esimerkkiratkaisuja kunhan pidän ajatuksen mukana.

Vertaisoppijat: Opiskelen mielelläni matematiikkaa yhdessä muiden opiskelijoiden kanssa ja laskiessani toivon, että saan neuvoa, jos en kykene itsenäisesti tehtävää ratkaisemaan. Kiinnitän huomiota esimerkkeihin ja koen, että matematiikan oppiminen on tarpeellista. Tehtäviä laskettaessa on mielestäni tärkeää saada oikea vastaus, vaikka joissakin kohdissa ratkaisua olisi virheitä. Pidän siitä, että yrittämisestä palkitaan.

Tukea tarvitsevat: Opiskellessani matematiikkaa haluan, että minua opastetaan henkilökohtaisesti vaikeissa kohdissa. Opettajan antamat esimerkit ja opetustapa vaikuttavat paljon siihen, miten omaksun asian. En mielelläni sovella malliratkaisuja uusiin tehtäviin. Jätän vaikeat tehtävät tekemättä tai kesken. Matematiikan "kieli" vaikuttaa minusta vaikealta.

Omin päin oppijat: Pystyn oppimaan matematiikkaa, jos koen tarvitsevani sitä. En laske tehtäviä mielelläni kavereiden kanssa, vaan opin parhaiten itsekseni pohtimalla. En myöskään tarvitse opettajan tukea oppimisessani. Laskutehtävien kopioiminen ei edistä oppimistani.

Osaajat: Haluan oppia matematiikkaa syvällisesti, enkä halua opetella asioita ulkoa. Laskiessani vaikeaa tehtävää en luovuta helpolla, vaan yritän ratkaista sen. Pärjään mielestäni hyvin matematiikassa.

Kuvataan seuraavaksi erilaiset opiskelijaprofiilit aikaisemman tutkimuksen [15] mukaisesti. Ensimmäinen ryhmä on pintasuuntautuneet mallista oppijat. Ryhmässä korostuu epävarmuus oman osaamisen suhteen, asenne ei ole kaikkein positiivisin ja oppiminen on pinnallista. Tehtävien teossa onnistutaan, jos katsotaan mallia esimerkkiratkaisusta. Oppimisen kannalta tällainen opiskelija jää vaille syvällistä oppimista vaikka onkin ollut valmis käyttämään keskimääräistä enemmänkin vaivaa oppimisensa eteen. Erilaisten mallien ulkoa opiskelu ei ole lopulta vähemmän työläämpi kuin syvällisemmän ymmärryksen tavoittelu ilman mallien opettelua.

Vertaisoppijoille yhteisöllisyys on tärkeää, ja heidän asenteensa matematiikan oppimiseen on positiivinen. Parhaat oppimistulokset koetaan saavutettavan, kun joku kaveriporukasta osaa neuvoa asian heille "kädestä pitäen". Lisäksi ryhmän opiskelijat kokevat heidän yrittämisensä huomioimisen innostavan opiskelun jatkamista. Oppimisen kannalta tuloksiin vaikuttaa opiskelijan vertaisryhmän rakenne, missä parhaassa tapauksessa syntyvä vertaiskeskustelu edesauttaa oppimista, mutta jossa riskinä on ainoastaan tehtävien kopioiminen ja sitä kautta suppea asioiden oppiminen.

Tukea tarvitsevien opiskelijoiden matematiikan osaaminen on kaikkein epävarminta. Heidän käsityksensä omasta osaamisestaan ja asenteensa matematiikan opiskelua kohtaan on huomattavan heikko. He luovuttavat helposti tehtäviä laskiessaan ja turvautuvat ulkoa opiskeluun, ja heille riittää saada tehtävät edes jotenkin valmiinnäköisiksi. Tukea tarvitseville matematiikan kieli on ongelmallista. Opiskelun onnistuminen koetaan olevan heistä itsestään riippumatonta ja siksi he toivovatkin jonkun neuvovan heitä “kädestä pitäen”. On selvää, että näistä lähtökohdista on erittäin vaikeata saada pitkäaikaisia onnistumisen elämyksiä ja sitä kautta motivaatiota matematiikan opiskeluun.

Omin päin opiskelevat kokevat oman osaamisensa positiivisena, eivätkä turvaudu kopioimiseen opettajalta tai vertaisopiskelijoilta. Jälkimmäisten merkitys opiskelussa ei myöskään korostu yhtä voimakkaasti kuin muissa ryhmissä. Ryhmällä on oppimisen kannalta hyvät mahdollisuudet omaksua tietoa heille ominaisella tavalla ja luoda syvällisempää ymmärrystä opiskeltavasta asiasta. Lähtökohtaisesti omin päin opiskeleville verkkopohjaiset itseopiskelumateriaalit, kuten matematiikkajumppa ja Math-Bridge ylipäätään, voivat olla mielekäs ja hyödyllinen tapa opiskella.

Osaajat ovat opiskelijoita, jotka eivät luovuta helpolla tehtäviä laskiessaan. Heillä on hyvä käsitys omasta osaamisestaan ja he kokevat osaavansa matematiikkaa. Asenne on positiivinen matematiikkaa kohtaan ja merkityssuuntautuneisuus on korostunut. He kokevat oppimisen ja matematiikassa onnistumisen riippuvan heistä itsestään toisin kuin tukea tarvitsevat. He kokevat oppivansa parhaiten, jos voivat käyttää luovaa päättelyä ja ottaa vastuun omasta oppimisestaan. Tällainen motivaatio ja halu oppia tuottaa ehdottomasti hyviä oppimistuloksia, mutta riskinä ovat väärät luulot omasta osaamisesta ja vääränlainen tapa tehdä töitä opiskelun eteen.

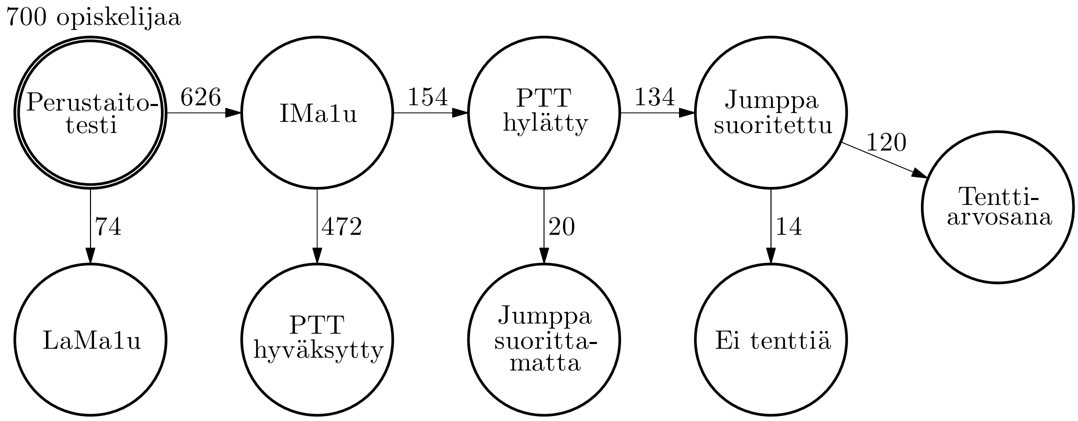
3. AINEISTO

Tässä diplomityössä tutkitaan niitä opiskelijoita, joiden matematiikan osaamisessa havaittiin puutteita matematiikan perustaitotestissä. Opiskelijoiden oppimista tutkittiin perustaitotestin lisäksi matematiikkajumpan, vapaaehtoisen uusintatestin sekä Insinöörimatematiikka 1 -kurssin menestyksen kautta. Lisäksi opiskelijoille teetettiin molempien testien yhteydessä opiskelijaprofiilikysely, jonka perusteella saatiin aiempaa tutkimusta hyödyntäen tietoa opiskelijoiden opiskelumenetelmistä, -asenteista sekä -motivaatiosta.

3.1 Kohdejoukko

Tampereen teknillisessä yliopistossa syksyllä 2011 opintonsa aloittaneet uudet opiskeljiat suorittivat matematiikan perustaitotestin. Kuten edellisessä luvussa todettiin, perustaitotestillä kartoitetaan ongelmia matematiikan perusasioiden osaamisessa. Testissä oli 16 tehtävää lukiomatematiikan eri aihealueista, ja kurssia Insinöörimatematiikka 1u suorittavien tuli saada tehtävistä vähintään seitsemän oikein saadakseen perustaitotestin hyväksytysti suoritettua. Vastaava raja Laaja matematiikka 1u -kurssia suorittaville oli 10 tehtävää. Opiskelijat saivat tehtävät tietokonepohjaisesti, mutta varsinainen laskeminen tapahtui kynällä ja paperilla. Palautus tapahtui myös tietokonepohjaisesti, ja ohjelma (Math-Bridge) antoi jokaisen tehtävän palautuksen kohdalla palautetta ratkaisun oikeellisuudesta. Opiskelijalla oli väärin menneen ratkaisun jälkeen mahdollisuus korjata vastaustaan ja palauttaa tehtävä uudelleen. Testi suoritettiin valvotuissa olosuhteissa, ja testiin käytettävää aikaa oli rajoitettu.

Perustaitotestin perusteella hylätyt opiskelijat ohjattiin matematiikkajumppaan, jonka hyväksytysti suorittaminen korvasi hylätyn perustaitotestin. Hyväksytysti suoritettu perustaitotesti tai matikkajumppa kuului osasuorituksena kursseille Insinöörimatematiikka 1u sekä Laaja matematiikka 1u. Vuonna 2011 perustaitotestin suoritti 700 opiskelijaa, joista 626 suoritti Insinöörimatematiikan kurssia ja loput Laajaa matematiikkaa. Insinöörimatematiikan opiskelijoista 154 ohjattiin matikkajumppaan. Näistä jumpan suoritti hyväksytysti 134 opiskelijaa. Tutkimusta varten saatiin Insinöörimatematiikka 1u:n ensimmäisen tentin tulokset ja tutkimus rajattiin koskemaan vain ensimmäiseen tenttiin osallistuneita opiskelijoita, jolloin jäljelle jäi 120 opiskelijaa. Näistä opiskelijoista poistettiin vielä kaksi epäselvien lokitietojen vuoksi,



Kuva 3.1: Tutkimusryhmän muodostuminen.

joten tutkimuksen kohdejoukoksi muodostui jäljelle jääneet 118 opiskelijaa. Kuvassa 3.1 on kuvattu opiskelijoiden jakautuminen perustaitotestin, suoritettavan kurssin, jumppasuorituksen sekä tenttiin osallistumisen perusteella.

3.2 Tutkimuskysymykset

Lähtökohtana tutkimukselle oli tutkia lokitietojen avulla opiskelijoiden toimintaa matematiikkajumpassa. Lokitiedoista pystytään opiskelijoiden toiminnan lisäksi tutkimaan myös jumpan tehtäviä. Kuten edellisessä luvussa kerrottiin, aikaisemmissa tutkimuksissa Tampereen teknillisellä yliopistolla [15] tutkittiin perustaitotestin, matematiikkajumpan ja kyselylomakkeiden avulla opiskelijoiden motivaatioita ja asenteita matematiikan opiskeluun ja tämän avulla määriteltiin opiskelijaprofiilit, joihin opiskelijat itsensä liittävät parhaaksi katsomallaan tavalla. Tämä tutkimus pyrkii toimimaan jatkumona aiemmalle tutkimukselle. Tutkimuksen avulla pyritään vastaamaan seuraaviin tutkimuskysymyksiin:

1. Käyttäytyvätkö erilaiset opiskelijaprofiilit eri tavoin matematiikkajumpassa?
2. Voidaanko lokitietojen perusteella ryhmitellä opiskelijoita?
3. Voidaanko kohdan 2 ryhmien perusteella löytää opiskelijat, joilla on ongelmia matematiikan opiskelussa?
4. Kenelle matematiikkajumppa on hyödyllinen?
5. Voidaanko lokitietojen perusteella havaita ongelmallisia tehtäviä ja tehtävätyyppejä?

3.3 Aineiston kerääminen

Tärkeimmät tiedot opiskelijoista kerättiin Math-Bridge-järjestelmän lokitiedoista. Lokitiedot järjestettiin opiskelijakohtaisesti aikajärjestykseen, jolloin niitä oli myöhemmin helppo käsitellä. Lokitiedoista saatiin jokainen opiskelijan tekemä tapahtuma Math-Bridge-järjestelmään kirjautumisen jälkeen. Lokitiedoista nähtiin ajallisesti opiskelijan tekemät tehtävät ja niiden ratkaisuyritykset.

Opiskelijoista saatiin myös tiedoksi perustaitotestin ja jälkitestin tulokset sekä näiden yhteydessä kysytyt opiskelijaprofiilit. Näiden avulla pystyttiin vertailemaan erilaisten opiskelijoiden toimintaa matematiikkajumpassa sekä tarkastelemaan matematiikkajumpan hyötyjä testien pistemäärän kehityksen perusteella.

Kolmas matematiikan osaamista testaava muuttuja saatiin niille opiskelijoille, jotka osallistuivat Insinöörimatematiikka 1u:n ensimmäiseen tenttiin syksyllä 2011. Tämän avulla voidaan tarkastella matematiikkajumppaan ohjattujen opiskelijoiden osaamista yliopistomatematiikan peruskurssilla sekä matematiikkajumpan hyödyllisyyttä näille opiskelijoille. Kuten aliluvussa 3.1 todettiin, niin kaikilta tutkimusryhmän opiskelijoilta löytyy kyseisen tentin arvosana.

3.4 Aineiston kuvailua

Alkuperäinen aineisto koostui 88 394 rivistä Math-Bridge-järjestelmän lokitietoja. Yksi rivi sisälsi yhden käyttäjän yhden tapahtuman, esimerkiksi tehtävän avaamisen. Lokitiedot koottiin sarkaineroteltuun tekstitiedostoon, jossa sarkaimilla eroteltiin sarakkeiksi lokitiedoista seuraavat tiedot:

- aikaleima (dd.mm.yyyy hh.mm.ss),
- opiskelijan nimi, intranet-tunnus ja opiskelijanumero,
- tapahtuman tyyppi,
  - `UserLoggedIn` (käyttäjä kirjautuu sisään)
  - `UserLoggedOut` (käyttäjä kirjautuu ulos)
  - `ExerciseStarted` (käyttäjä avasi tehtävän)
  - `ExerciseStep` (käyttäjä yritti ratkaista tehtävän tai katsoi mallivastausta)
  - `PagePresented` (käyttäjä katsoi jumppakirjan sivua)
- IP-osoite,
- käytetty selain,
- jumppakirjan tyyppi,
- jumppakirjan sivunumero,
- tehtävän koodi,
- tehtävän ratkaisuun käytetty aika millisekunteina tehtävän avaamisesta,
- ratkaisuyrityksen palaute,

  - `1.0`(tehtävä ratkaistiin oikein)
  - `0.0`(tehtävä ratkaistiin väärin)
  - `-1.0`(katsottiin tehtävän malliratkaisua)
- opiskelijan ratkaisu sekä
- kirjautumisen kokonaiskesto.

Aineistoon tutustuessa havaittiin nopeasti, että kaikki sarakkeet eivät ole tarpeellisia. Erityisesti jouduttiin luopumaan tehtävien ratkaisemiseen käytetystä ajasta sekä kirjautumisen kokonaiskestosta, sillä näiden avulla ei saatu riittävän tarkasti mitattua opiskelijan todellista aikaa järjestelmässä tai yksittäisen tehtävän kohdalla. Lokitiedoista saatiin ratkaisun kokonaisaika, jos käyttäjä oli yrittänyt ratkaista tehtävän, mutta malliratkaisun katsomiseen kulunutta aikaa ei ollut mahdollista saada. Lisäksi käyttäjän kokonaisaika kirjautuneena saattoi sisältää ylimääräistä aikaa esimerkiksi tilanteessa, jossa käyttäjä ei ollut tehnyt mitään 30 minuuttiin, jolloin järjestelmä kirjasi käyttäjän automaattisesti ulos. Turhien sarakkeiden lisäksi havaittiin, että aineistosta voidaan poistaa myös sellaiset rivit, joita tutkimuksessa ei voida hyödyntää, kuten sisään- ja uloskirjautuminen sekä oppikirjan sivujen selaaminen.

Sarkaineroteltu tekstitiedosto saatiin avattua helpompaa muokkaamista ja hahmottamista varten Microsoft Officen Excel-ohjelmalla. Aluksi datasta poistettiin ylimääräiset sarakkeet ja loput sarakkeet siistittiin tulevaa käyttötarkoitusta varten selkeämpään muotoon. Jäljelle jääneisiin sarakkeisiin valikoitui tiedot:

- aikaleima (dd.mm.yyyy hh.mm.ss),
- opiskelijanumero,
- tapahtuman tyyppi (ExerciseStarted tai ExerciseStep),
- tehtävän koodi sekä
- ratkaisun palaute.

Näiden lisäksi lokitietoihin lisättiin käsin oppilaskohtaisia tietoja:

- koulutusohjelma,
- perustaitotestin tulos,
- opiskelijaprofiilikyselyn tulos,
- jumpan jälkeen uusitun perustaitotestin tulos,
- jumpan jälkeen uusitun opiskelijaprofiilikyselyn tulos,
- ensimmäisen tentin arvosana kurssista Insinöörimatematiikka 1u sekä
- sukupuoli.

Liitteessä 2.3 on esitetty esimerkki lokitiedoista. Tehtävien tunnistuksessa käytetyt koodit perustuivat tehtävien tyyppiin. Tehtäviä oli yhteensä 71 kappaletta ja ne oli jaettu kymmeneen eri kategoriaan. Tarkemmat tehtävien esittelyt löytyvät liitteestä C. Tehtävätyypeistä käytettiin lyhenteitä, jotka esitellään lyhenteiden mukaisessa aakkosjärjestyksessä seuraavaksi.

Derivointi (der): Yhdeksän derivointitehtävää, joissa derivoidaan polynomifunktioita, trigonometrisia funktioita, yhdistettyjä funktoita sekä logaritmifunktioita.

Yhtälöt (eq): Seitsemän yhtälönratkaisutehtävää, joissa ratkaistaan ensimmäisen, toisen ja kolmannen asteen yhtälöitä, itseisarvo-, neliöjuuri- sekä murtoyhtälöitä.

Lausekkeet (exp): Kuusi tehtävää, joissa lasketaan annetujen funktioiden sekä yhdistettyjen funktioiden arvoa, sievennetään lausekkeita sekä murtolausekkeita ja määritetään käänteisfunktioita.

Epäyhtälöt (ieq): Seitsemän epäyhtälötehtävää, joissa ratkaistaan ensimmäisen ja toisen asteen epäyhtälöitä, itseisarvo- sekä murtoepäyhtälöitä ja näiden sekotuksia.

Integrointi (int): Kahdeksan integrointitehtävää, joissa ratkaistaan polynomi- ja eksponentti- sekä itsearvofunktioiden integraalifunktioita sekä määrättyjä integraaleja.

Raja-arvot (lim): Kuusi raja-arvotehtävää, joissa määritetään polynomien, murtolausekkeide, trigonometristen funktioiden, vakioiden sekä paloittain määriteltyjen funktioiden raja-arvoja.

Logaritmit (log): Seitsemän logaritmitehtävää, joissa ratkaistaan eksponenttiyhtälöitä, logaritmiyhtälöitä sekä -epäyhtälöitä erilaisilla logaritmin kannoilla.

Lukuarvot (n): Seitsemän tehtävää, joissa testataan käsitteiden käänteisluku, vastaluku sekä itseisarvo osaamista. Lisäksi tehtävissä lasketaan lausekkeiden arvoja annetuilla parametreilla.

Potenssit (pow): Seitsemän tehtävää, joissa lasketaan lausekkeen arvoja, sievennetään tai etsitään juuria potenssi ja neliöjuurilausekkeille.

Trigonometria (tri): Seitsemän trigonometrian tehtävää, joissa on tehtäviä asteista ja radiaaneista sekä trigonometrisista yhtälöistä käyttäen siniä, kosinia ja tangenttia.

3.5 Aineiston käsittely

Halutulla tavalla jäsennelty lokitietoaineisto haluttiin saada koostettua sellaiseen muotoon, jossa nähtäisiin kootusti yksittäisen opiskelijan toiminta jumpassa. Tätä varten ohjelmoitiin Pythonilla ohjelma, joka luki lokitiedoston ja laski jokaiselle opiskelijalle halutut tiedot. Python on ohjelmointikieli, joka sopii skirptiparadigman ohjelmointikielenä hyvin kyseisenlaiseen tekstitiedoston käsittelyyn. Ohjelman avulla koostettua aineistoa kutsutaan jatkossa datamatriisiksi, jonka jokaisella rivillä on yhden opiskelijan tiedot ja sarakkeet sisältävät seuraavat tiedot:

- opiskelijanumero,
- jokaiselle tehtävälle viisi saraketta, jotka ovat
  - tehtävän avauskerrat,
  - tehtävän ratkaisuyritykset,
  - mallivastauksen katsomisen lukumäärä,
  - aikaleima käyttäjän ensimmäisestä tehtävän avauskerrasta ja
  - aikaleima käyttäjän ensimmäisestä oikeasta kyseisen tehtävän ratkaisusta,
- koulutusohjelma,
- perustaitotesti ennen jumppaa,
- opiskelijaprofiilikysely ennen jumppaa,
- perustaitotesti jumpan jälkeen,
- opiskelijaprofiilikysely jumpan jälkeen,
- Insinöörimatematiikka 1u:n ensimmäisen tentin arvosana,
- sukupuoli,
- avauskertojen kokonaismäärä,
- yrityskertojen kokonaismäärä sekä
- mallivastausten katsomisen kokonaismäärä.

Tiedosto tallennettiin sarkainerotelluksi tekstitiedostoksi. Tämän jälkeen tutkimuksen edetessä lisättiin vielä viisi uutta saraketta:

- avauskertojen ja mallivastauksen katsomisten erotus,
- parannus perustaitotestien välillä,
- kokonaisklikkausten määrä,
- oikeiden vastausten lukumäärä sekä
- väärien vastausten lukumäärä.

4. TUTKIMUSMENETELMÄT

Tässä tutkimuksessa käytetään pääasiallisena tutkimusmenetelmänä tilastotieteen tarjoamia työkaluja. Tilastotiede voidaan jakaa tutkimuksen kannalta kahteen osaan: kuvailevaan (descriptive) sekä päättelevään (inferential) tilastotieteeseen. Ensimmäistä käytetään tutkimuksen alussa, jotta saataisiin jonkinlainen yhteenveto mitatusta datasta. Jälkimmäisen tarkoituksena on pystyä päättelemään aineistosta johtopäätöksiä riittävän luotettavasti.

Tässä luvussa esiteltävät menetelmät ovat peräisin useasta eri lähteestä. Tärkeimpinä lähteinä tilastotieteen perustyökalujen tapauksessa ovat olleet Walpolen et al. [17] sekä Metsämuurosen [13] teokset. Näiden lisäki on käytetty Gorardin [6], Johnsonin ja Wichhernin [8] sekä Miltonin ja Arnoldin [14] teoksia. Erityisesti klusterianalyysin kohdalla lähteinä on käytetty Johnsonin ja Wichernin lisäksi Jainin et al. [7], Kriegelin et al. [9] sekä Xun et al. artikkeleita. Normaalijakauman testauksen taulukkoarvot saatiin Masseyn [12] sekä Lillieforsin [11] artikkeleista.

4.1 Aineiston kuvaaminen

Tilastotieteessä tilastointia eli mittaamista koskeva tilastollinen perusyksikkö on *tilastoyksikkö*, esimerkiksi opiskelija tai muu tietoa tuottava yksikkö, jolta tiedot kerätään. Tilastoyksiköstä käytetään myös nimitystä *havainto*. Tilastoyksiköt muodostavat tutkimuksen *otoksen* (esimerkiksi lukiolaiset) ja koko tilastoyksiköiden joukko (kaikki opiskelijat) muodostavat *populaation*. Tilastoyksiköltä kerätään tietoa jostakin *muuttujasta*, esimerkiksi opintopisteiden määrästä tai tentin arvosanasta. Keräämällä muuttujien tiedot saamme havaintoaineiston $x_1, x_2, ..., x_n$, missä x_i käsittää yhden muuttujan arvon eli *havaintoarvon*. Useamman tilastoyksikön tapauksessa x_i käsittää haviantoarvot kaikille tilastoyksikölle, jolloin se on siis *vektori*. Tässä tutkielmassa vektoria käsitellään lähinnä listana lukuja, joten sen tarkempi matemaattinen määrittely sivuutetaan tässä tutkielmassa. Esimerkkinä taulukossa 4.1 on esitetty yhdentoista kuvitteellisen opiskelijan tietoja eräältä kurssilta.

Tällä tavoin esitettyä taulukkoa kutsutaan myös *datamatriisiksi*. Opiskelijat ovat esimerkkiaineiston tilastoyksiköitä. Opiskelijoista on kerätty tiedot koulutusohjelmasta, vuosikurssista, harjoituspisteistä sekä tentin arvosana joltakin kurssilta. Nämä muuttujat ovat siis yhdentoista havaintoarvon suuruisia vektoreita.

Taulukko 4.1: Yhdentoista opiskelijan esimerkkidata.

Opiskelija-numero	Koulutus-ohjelma	Vuosi-kurssi	Harjoitus-pisteet	Tentin arvosana
9551	Sähkötekniikka	2	37	5
3080	Konetekniikka	2	33	3
6461	Biotekniikka	6	48	4
5373	Tietotekniikka	3	31	3
9021	Tuotantotalous	2	21	2
7858	Konetekniikka	3	29	3
5108	Sähkötekniikka	4	42	4
1166	Tietotekniikka	5	28	2
8392	Konetekniikka	2	29	1
5002	Tietotekniikka	4	47	3
6749	Tuotantotalous	3	19	0

4.1.1 Mitta-asteikko

Muuttujia käsiteltäessä ja niille laskettaessa erilaisia tunnuslukuja on huomioitava millaisella *mitta-asteikolla* muuttuja on mitattu ja miten niitä kuvataan. Mitta-asteikkoja on yleisesti viisi erilaista: luokitteluasteikko, järjestysasteikko, intervalliasteikko, suhdeasteikko sekä absoluuttinen asteikko.

Luokitteluasteikko: Luokitteluasteikosta käytetään myös nimiä laatuero- tai nominaaliasteikko (Nominal Scale). Sillä voidaan luokitella tilastoyksiköitä muuttujan suhteen, joka kuvaa laatua, ei määrää. Luokitteluasteikolla mitattavan muuttujan havaintoarvoja ei voida järjestää suuruus- tai paremmuusjärjestykseen. Tällainen muuttuja voi olla esimerkiksi silmien väri tai sukupuoli.

Järjestysasteikko: Järjestysasteikolla eli ordinaaliasteikolla (Ordinal Scale) kuvataan yksinkertaisinta järjestystä, jossa muuttujan havaintoarvot voidaan järjestää siihen liittyvän ominaisuuden perusteella järjestykseen. Eri havaintoarvojen välistä keskinäistä suhdetta ei voida kuitenkaan määrittää. Tällaisia muuttujia ovat esimerkiksi armeijan arvojärjestys tai koulutustausta.

Välimatka-asteikko: Välimatka- eli intervalliasteikolla (Interval Scale) saadaan havaintoarvojen välille tietty etäisyys ja suhde. Yleensä mittauksissa käytetään vähintäänkin välimatka-asteikkoa, kun on kysymys jonkin ilmiön tai testin havainnoista ja realisoituneista arvoista. Esimerkkinä välimatka-asteikolla mitattavasta muuttujasta on lämpötila, jolle ei ole olemassa yksikäsitteistä nollakohtaa kaikilla mitta-asteikoilla, esimerkiksi Celsius- ja Kelvin-asteikolla nollakohta ei vastaa samaa lämpötilaa.

Suhdeasteikko: Suhdeasteikko (Ratio Scale) on tarkennus välimatka-asteikosta.

Tällöin voidaan yksikäsitteisesti ilmaista mitattavan ominaisuuden loppuminen. Esimerkkejä suhdeasteikosta ovat pituus ja paino.

Absoluuttinen asteikko: Absoluuttinen asteikko (Absolute Scale) on tarkennus suhdeasteikosta, missä saadulla havaintoarvolla on vain yksi merkitys ja yksikkö. Esimerkiksi tentissä saadut pisteet tai laskuharjoituspisteet, kuten muutkin lukumäärää kuvaavat muuttujat voidaan kuvata absoluuttisella asteikolla.

Tarkastellaan esimerkkinä esimerkkiaineistoa taulukossa 4.1. Taulukon muuttujista koulutusohjelmaa mitataan luokitteluasteikolla, vuosikurssia järjestysasteikolla ja harjoituspisteitä suhdeasteikolla. Tentin arvosanaa voidaan mitata tilanteesta riippuen eri mitta-asteikoilla. Vähimmillään tenttimenestystä mitataan järjestysasteikolla, jos halutaan korostaa arvosanan muodostumista tehtävistä saaduista yhteispisteistä, jolloin samatkin arvosanat voivat olla eriarvoisia. Toisaalta tutkimuksen kannalta on mielekästä käyttää pistemäärää omana muuttujanaan, jos sellainen on saatavissa, jolloin tentin arvosanaa voidaan mitata absoluuttisella asteikolla.

Aineistoa käsiteltäessä tietokoneella tulee ottaa huomioon havaintoaineiston alkioiden numerokoodaus. Esimerkiksi sukupuolta käsiteltäessä tulee alkiot "Mies" ja "Nainen" koodata esimerkiksi numeroarvoilla 1 ja 2.

4.1.2 Frekvenssi

Mitta-asteikon valinnan jälkeen aineiston kuvaaminen voidaan aloittaa laskemalla muuttujien eri havaintoarvojen esiintymisten lukumäärä. Tätä kutsutaan *frekvenssiksi*, josta voidaan käyttää myös nimiä esiintymistiheys tai -taajuus. Frekvenssi voidaan esittää joko suoraan lukumääränä tai suhteellisena prosenttiosuutena koko aineistosta. Tutkitaan esimerkiksi taulukon 4.1 opiskelijoita ja tentin arvosanaa. Erilaisille tentin arvosanoille saadaan taulukon 4.2 mukaiset frekvenssit f sekä prosenttiosuudet $f\%$.

Taulukko 4.2: Arvosanojen jakautuminen.

Arvosana	f	$f\%$
0	1	9,1
1	1	9,1
2	2	18,2
3	4	36,4
4	2	18,2
5	1	9,1

Suurella aineistolla taulukkoa havainnollistavampi tapa esittää frekvenssit on esittää ne *histogrammina*. Histogrammi eli pylväsdiagrammi kuvaa eri muuttujan havaintoarvot omina pylväinään ja pylvään korkeus kertoo kyseisen arvon frekvenssin.

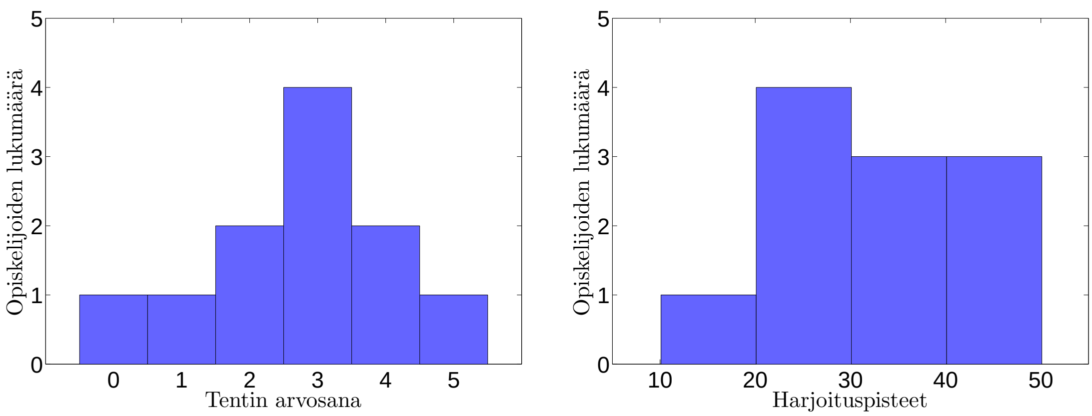
Histogrammia varten täytyy määrittää muuttujan mahdollisten ja tutkimuksessa esiintyneiden havaintoarvojen rajat. Yksinkertainen tapa kuvata aineiston rajoja on otoksen *vaihteluväli* $[x_{min}, x_{max}]$ tai yhdellä luvulla ilmaistuna välin rajojen erotus $x_{max} - x_{min}$.

Jos muuttujan arvojen vaihteluväli on suuri verrattuna tilastoalkioiden lukumäärään, ei jokaiselle havaintoarvolle laskettu frekvenssi kuvaa aineistoa mielekkäästi. Tällöin aineistoa *tiivistetään* luokittelemalla eli ryhmittelemällä sitä sopivan kokoisiin luokkiin. Esimerkiksi taulukossa 4.1 oleva muuttuja "Harjoituspisteet" voitaisiin luokitella esimerkiksi taulukossa 4.3 esitetyllä tavalla.

Taulukko 4.3: Harjoituspisteiden muodostamat luokat ja niiden jakautuminen.

Luokka	f	f%
11–20	1	18,2
21–30	4	45,5
31–40	3	18,2
41–50	3	18,2

Tällä tavoin saadaan varsinkin suuremmalla aineistolla järkevämpi tapa havainnollistaa aineistoa. Kuvataan seuraavaksi taulukoiden 4.2 ja 4.3 histogrammit.



Kuva 4.1: Esimerkkiaineiston histogrammit tentin arvosanan sekä harjoituspisteiden perusteella.

Kuvasta 4.1 nähdään taulukkoa nopeammin suurimmat frekvenssit ja voidaan suunnitella tulevaa tutkimusta aineiston pohjalta. Frekvenssien pohjalta voidaan saada alustava näkemys aineiston keskimääräisistä arvoista sekä vaihtelevuudesta, jotka määritellään seuraavaksi.

4.1.3 Keski- ja hajontaluvut

Mittausaineiston kuvaamisen käytetyimpiä tunnuslukuja ovat erilaiset keski- ja hajontaluvut. Näillä saadaan karkea kuva mitatusta otoksesta tietyn muuttujan suhteen. Tunnetuin tällainen tunnusluku on *keskiarvo* (mean), jonka avulla saadaan kuvattua keskimääräistä muuttujan arvoa koko havaintoaineistosta. Tarkasti määriteltynä kyse on otoksen *aritmeettisesta keskiarvosta* (arithmetic mean), joka määritellään seuraavasti.

Määritelmä 4.1.1. Olkoon otoksesta mitattu muuttuja x ja sen havaintoarvot $x_1, x_2, ..., x_n$. Muuttujan *(otos)keskiarvo* $\bar{x}$ on

$$\bar{x} = \sum_{i=1}^{n} \frac{x_i}{n} = \frac{x_1 + x_2 + \cdots + x_n}{n}.$$

Keskiarvo saadaan laskemalla muuttujan havaintoarvot yhteen ja jakamalla summa havaintoarvojen lukumäärällä. Keskiarvo voidaan laskea vain sellaiselle muuttujalle, jota mitataan vähintään välimatka-asteikolla.

Tapauksissa, joissa muuttujaa mitataan enintään järjestysasteikolla, käytetään keskiarvon sijasta keskilukuna *mediaania*. Mediaani saadaan järjestämällä havaintoaineisto kasvavaan järjestykseen ja valitsemalla aineistosta keskimmäinen havaintoarvo. Jos havaintoaineiston koko on parillinen, niin mediaani on kahden keskimmäisimmän havaintoarvon keskiarvo tai vaihtoehtoisesti ilmoitetaan molemmat havaintoarvot. Mediaani on siis se arvo, jota suurempia ja pienempiä on 50 % aineistosta.

Luokittelumuuttujalle ei voida laskea mediaania, koska sitä ei voida yksikäsitteisesti järjestää mihinkään järjestykseen. Tällöin keskilukuna käytetään *moodia*. Otoksen moodi on se luokittelumuuttujan havaintoarvo, jota otoksessa on eniten.

Keskiarvo kertoo siis keskimääräisen arvon mitatusta havaintoaineistosta. Sen lisäksi aineistoa kuvataan erilaisilla hajontaluvuilla, joiden avulla saadaan käsitys aineiston havaintoarvojen poikkeamasta suhteessa keskiarvoon. Intuitiivisesti hajonta on havaintoarvojen keskimääräistä poikkeamaa keskiarvosta. Tätä kutsutaan *absoluuttiseksi hajonnaksi* tai *itseispoikkeamaksi* (Absolute deviation), joka määritellään seuraavasti.

Määritelmä 4.1.2. Muuttujan x havaintoarvojen $x_1, x_2, ..., x_n$ *absoluuttinen hajonta*, d on

$$d = \sum_{i=1}^{n} \frac{|x_i - \bar{x}|}{n} = \frac{|x_1 - \bar{x}| + |x_2 - \bar{x}| + \cdots + |x_n - \bar{x}|}{n}.$$

Absoluuttinen hajonta on siis havaintoarvojen ja keskiarvon etäisyyksien keskiarvo. Tätä muotoa hajonnasta ei kuitenkaan yleensä käytetä, vaan hajonta määritellään *varianssin* avulla, jolloin on kysymys *keskihajonnasta*. Varianssi on käytetyin hajontaluku tilastotieteessä. Se saadaan laskemalla keskiarvo havaintoarvojen ja keskiarvon etäisyyksien neliöstä. Keskihajonta saadaan varianssin neliöjuurena.

Määritellään seuraavaksi varianssi ja keskihajonta.

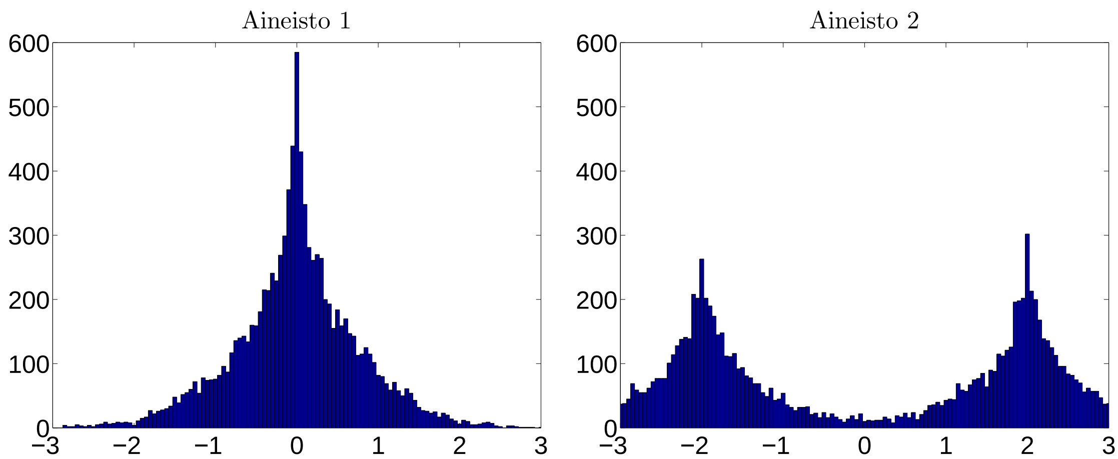
Määritelmä 4.1.3. Muuttujan x havaintoarvojen $x_1, x_2, ..., x_n$ *varianssi*, s^2 on

$$s^2 = \sum_{i=1}^{n} \frac{(x_i - \bar{x})^2}{n-1}.$$

Varianssin s^2 positiivista neliöjuurta kutsutaan otoksen *keskihajonnaksi*, s, joka saadaan siis

$$s = \sqrt{s^2}.$$

Pelkkiä keskiarvoja vertailemalla ei nähdä kahden aineiston välistä eroa ja sen vuoksi keskilukujen lisäksi käytetään hajontalukuja varianssi ja keskihajonta. Varianssia käytetään lisäksi muiden tilastollisten tunnuslukujen laskennassa ja tilastollisissa menetelmissä. Kuvassa 4.2 on esitetty esimerkki kahdesta satunnaisesti tuotetusta aineistosta, joiden molempien keskiarvo on nolla, mutta varianssit saavat eri arvot.



Kuva 4.2: Esimerkki aineistoista, joilla on sama keskiarvo mutta erilainen hajonta.

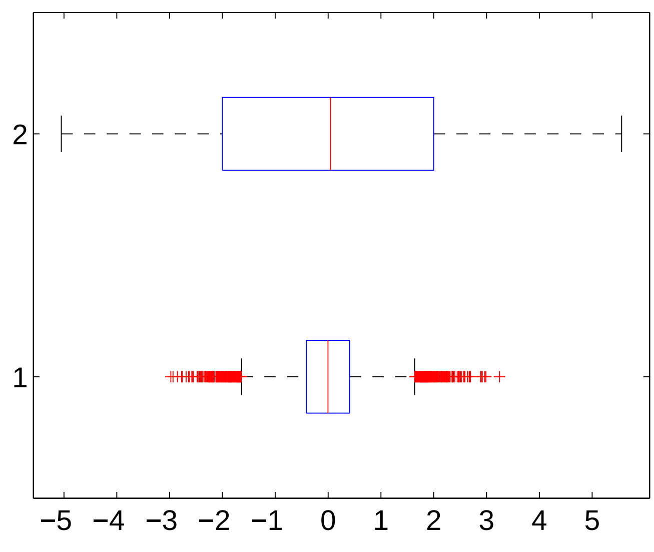
Aineiston 1 varianssi on luokkaa 0,6 ja aineiston 2 varianssi puolestaan luokkaa 4,6. Hajontaluvuilla on tärkeä rooli realistisen tutkimusaineiston kuvaamisessa. Aineistosta tehdyissä päätelmissä hajontaluvut on otettava huomioon varsinkin tilanteissa, joissa verrataan keskenään useampia kuin yhtä aineistoa, jotka on tuotettu mittaamalla samaa ilmiötä esimerkiksi eri tutkimusryhmistä. Keskiarvon ja varianssin avulla voidaan myös tutkia muuttujan havaintoarvojen jakautumista erilaisiin todennäköisyysjakaumiin, joista kerrotaan lisää aliluvussa 4.2.1.

4.1.4 Kvantiilit

Hajontalukujen lisäksi muuttujan jakautumista voidaan kuvata *kvantiilien* avulla. Muuttujan kvantiili, $q(f)$, on arvo, jota pienempiä tai yhtäsuuria on f prosenttia muuttujan arvoista. Lukua f kutsutaan *prosenttipisteeksi*. Yksinkertaisin kvantiili on aiemmin määritelty mediaani, joka on järjestetyn muuttujan keskimmäinen havaintoarvo, eli jakaa muuttujan arvot puoliksi. Mediaani on siis $q(0.5)$, missä prosenttipiste $f = 0.5$ eli 50 %.

Mediaanin lisäksi käytetään yleisesti 25 % ja 75 % prosenttipisteitä, joita kutsutaan *kvartiileiksi* ja erityisesti $q(0.25)$ on *alakvartiili* ja $q(0.75)$ on *yläkvartiili*. Muita yleisiä kvantiileja on *desiilit* $q(0.1)$, $q(0.2)$, $q(0.3)$,...,$q(0.9)$ ja *sentiilit* $q(0.01)$, $q(0.02)$, $q(0.03)$,...,$q(0.99)$. Erityisesti kvantiilien eroa kutsutaan *kvantiiliväliksi*, esimerkiksi $q(0.75) - q(0.25)$ on kvartiiliväli.

Kvantiilit ovat siis hyödyllinen tapa kuvata aineiston vaihteluvälejä sekä jakaa ja luokitella aineistoa ryhmiin. Kvantiileja käytetäänkin esimerkiksi *laatikkokuvaajissa* (Boxplot tai Box-and-Whisker Plot). Laatikkokuvaajaan on merkitty laatikko, jonka vasen ja oikea reuna ovat aineiston kvartiilit $q(0.25)$ ja $q(0.75)$, jolloin laatikko kuvaa *sisäkvartiiliväliä* (interquartile range). Laatikon keskellä on lisäksi viiva mediaania $q(0.5)$ varten. Laatikkokuvaajassa on lisäksi viikset, joiden tarkoituksena on muodostua sisäkvartiilivälin ulkopuolisten havaintojen mukaisesti. Viikeset piirretään tyypillisesti kuitenkin kattamaan vain ne pisteet, jotka sijaitsevat enintään 1,5 kertaa laatikon leveyden etäisyydellä laatikon reunasta. Tämän vuoksi viiksien ulkopuolelle piirretäänkin erillisinä pisteinä näistä rajoista ulos jääneet, poikkeavat havainnot (outliers). Esimerkki laatikkokuvaajasta on esitetty kuvassa 4.3.



Kuva 4.3: Esimerkkiaineistojen 1 ja 2 (Kuva 4.2) laatikkokuvaajat.

Kuvassa on käytetty samaa esimerkkiaineistoa kuin erilaisia variansseja havainnollistavassa kuvassa 4.2. Kuvasta nähdään aineistojen keskiarvojen olevan nolla, mutta sisäkvartiilivälit ovat selvästi eri suuret. Aineisto 1 koostuu yhdestä terävästä piikistä keskiarvon ympärillä. Laatikkokuvaajassa kyseisen aineiston laatikko on hyvin kapea, koska aineisto on niin tiiviisti keskiarvon läheisyydessä. Toisaalta kuvaajan viikset eivät ole riittäneet kuvaamaan selvästi keskiarvosta poikkeavia arvoja, jolloin on muodostunut paljon +-merkillä kuvattuja yksittäisiä poikkeavia arvoja. Toisen aineiston kohdalla aineisto koostui kahdesta piikistä, jolloin keskiarvon läheisyydessä ei ollut paljoa havaintoja. Tämän vuoksi kvartiiliväli on suurempi, jolloin laatikko on leveämpi kuin ensimmäisen aineiston kohdalla. Kuten aikaisemmin todettiin, laatikon leveys vaikuttaa viiksien leveyteen ja näin ollen viikset kattavat kaikki sisäkvartiilivälin ulkopuolelle jääneet havainnot.

Kvantiilien ja laatikkokuvaajien avulla ei voida tehdä tilastollisia johtopäätöksiä, vaan niiden avulla voidaan kuvata aineistoa helposti ja nopeasti sekä verrata sitä mahdollisesti toiseen, samaa populaatiota mittaavaan tutkimusaineistoon. Laatikkokuvaajasta on mahdollista havaita selvästi muusta otoksesta poikkeavat havaintoarvot. Luotettavuuden vuoksi tulisi kuitenkin käyttää erityisesti poikkeamien havaitsemiseen tarkoitettuja tilastollisia testejä.

4.2 Tilastollinen päättely

Aineiston kuvaamisella pyrittiin saamaan hyvä yleiskuva mitatusta aineistosta. Kuvaamisen jälkeen aineisto halutaan tutkia ja tulkita, esimerkiksi vertailemalla muuttujia ja tilastoyksiköitä eli havaintoja toisiinsa. Tutkimuksen tarkoituksena on saada vastauksia ennalta asetettuihin tutkimuskysymyksiin. Tässä aliluvussa esitellään tilastollisen päättelyn perusteita ja kaksi tutkimuksessa käytettyä tilastollista menetelmää aineiston tutkimiseen.

Tilastollinen päättely voidaan jakaa karkeasti kahteen eri menetelmään: induktiiviseen ja deduktiiviseen päättelyyn. Induktiivinen päättely tarkoittaa yksinkertaistettuna sitä, että yksittäisistä tapauksista pyritään päättelemään yleisiä lakeja tai yleistyksiä. Jos taas käytössä on jo jokin valmis teoria, jonka avulla voidaan pyrkiä ennustamaan yksittäisten tapausten toteutumista, niin puhutaan deduktiivisesta päättelystä. Todellisessa tutkimuksessa raja ei ole näin selkeä, vaan molempia menetelmiä käytetään tutkimuksen eri vaiheissa. Lisäksi tilastotiede hyödyntää myös vahvasti todennäköisyyden käsitettä, jota käytetään joko ennalta tiedettynä *etukäteistodennäköisyytenä* tai mitattuna *jälkikäteistodennäköisyytenä*.

4.2.1 Otosjakaumat

Tässä aliluvussa esitellään kaksi tutkimuksessa käytettyä todennäköisyysjakaumaa: normaalijakauma sekä t-jakauma. Lisäksi määritetään niihin liittyvä satunnaismuuttujan käsite. Tämän jälkeen esitellään myös menetelmä, jonka avulla voidaan tutkia muuttujan sopivuutta normaalijakaumaan. Näiden käsitteiden ja testien avulla voidaan valita tutkimukseen parhaiten sopivat menetelmät, kun muuttujien jakaumat tunnetaan.

Normaali- ja t-jakauma

Tilastotieteessä käytetään useita erilaisia jakaumia kuvaamaan otoksen jakautumista esimerkiksi jonkin muuttujan keskiarvon ja varianssin suhteen. Jakaumien avulla aineistolle saadaan tilastollinen malli eli todennäköisyysjakauma, joka kertoo muuttujan arvojen esiintymistodennäköisyyden. Ennen jakaumien esittelyä täytyy kuitenkin määritellä *satunnaismuuttujan* käsite. Satunnaismuuttuja kertoo mitattavalle muuttujalle sen havaintoarvoja vastaavan esiintymistodennäköisyyden jonkin todennäköisyysjakauman avulla. Toisin sanoen jos jokin mitattu muuttuja on jakautunut jonkin todennäköisyysjakauman mukaisesti, niin sitä voidaan mallintaa kyseisellä todennäköisyysjakaumalla ja käyttää mitattujen muuttujien sijaan satunnaismuuttujia. Tilastotieteessä erotetaan otoksesta mitattu reaalimaailman muuttuja ja satunnaismuuttuja toisistaan. Mitattuja muuttujia verrataan usein satunnaismuuttujien todennäköisyysjakaumaan ja niille voidaan tehdä tilastollisia testejä, jotka kertovat muuttujan mahdollisesta sopivuudesta johonkin todennäköisyysjakaumaan. Satunnaismuuttujien kohdalla käytetään keskiarvosta merkintää μ ja varianssista merkintää σ^2.

Yleisimpänä ja tärkeimpänä näistä todennäköisyysjakaumista on *normaalijakauma*, jonka mukaisesti muuttujien oletetaan käyttäytyvän monissa tilastollisissa menetelmissä. Määritellään normaalijakauman tiheysfunktio.

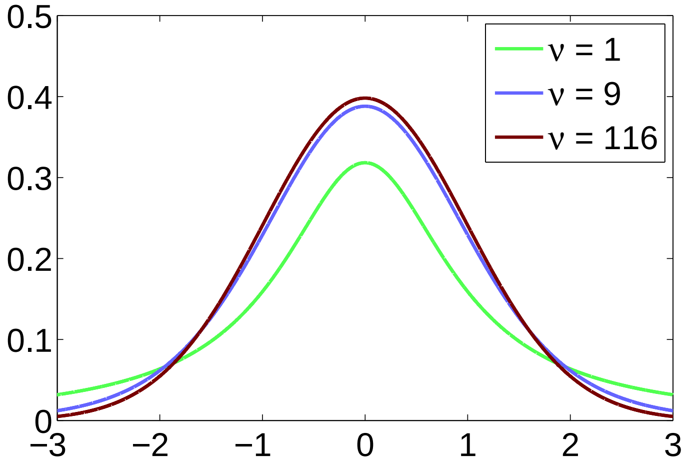
Määritelmä 4.2.1. Normaalijakautuneen satunnaismuuttujan X tiheysfunktio keskiarvolla μ ja varianssilla σ^2

$$n(X;\mu,\sigma^2) = \frac{1}{\sqrt{2\pi}\sigma}\mathrm{e}^{-\frac{1}{2\sigma^2}(X-\mu)^2},$$

missä $-\infty < X < \infty$.

Toinen tässä tutkimuksessa hyödyllinen todennäköisyysjakauma on normaalijakaumaan pohjautuva *t-jakauma*. Otoskoon ollessa pieni (alle 30) suositellaan normaalijakauman sijaan käytettäväksi t-jakaumaa, jolloin saadaan normaalijakaumaa tarkempia tuloksia. Jakauman ottaa huomioon otoskoon käyttämällä niin sanottuja *vapausasteita*. Suuremmillakin aineistoilla voidaan käyttää t-jakaumaa, koska otoskoon kasvaessa kasvavat myös vapausasteet ja suurilla vapausasteilla t-jakauma on oleellisesti sama kuin normaalijakauma. Määritellään seuraavaksi t-jakauman tiheysfunktio, josta on esimerkki kuvassa 4.4.

Määritelmä 4.2.2. t-jakautuneen satunnaismuuttujan X tiheysfunktio $f(X)$ va-



Kuva 4.4: t-jakauma kolmella eri vapausasteella. Vapausasteilla 116 jakauma on oleellisesti sama kuin normaalijakauma keskiarvolla $\mu = 0$ ja varianssilla $\sigma^2 = 1$.

pausasteilla ν on

$$f(X) = \frac{\Gamma\big((\nu+1)/2\big)}{\sqrt{\nu\pi}\Gamma(\nu/2)} \left(1 + \frac{X^2}{\nu}\right)^{(\nu+1)/2},$$

missä Γ on *gammafunktio*, joka määrätynä integraalina merkittynä on muotoa

$$\Gamma(r) = \int_0^\infty x^{r-1} \mathrm{e}^{-x} dx.$$

Näiden jakaumien lisäksi tilastotieteessä on myös muita paljon käytettyjä jakaumia kuten χ^2-, F- ja *Poissonjakauma*. Näillä on omat käyttötarkoituksensa, mutta niitä ei kuitenkaan tässä tutkielmassa esitellä tarkemmin.

Kolmogorovin–Smirnovin yhden otoksen testi

Havaitun jakauman sopivuutta, esimerkiksi normaalijakaumaan, voidaan testata *Kolmogorovin–Smirnovin yhden otoksen testillä*. Testin avulla voidaan testata sopivuutta myös muihin jakaumiin riippuen ilmiöstä otoksen ja muuttujan taustalla. Testin tuloksen perusteella voidaan päättää sopivat menetelmät muuttujien myöhempään analyysiin. Jotkin menetelmät olettavat muuttujilta normaalijakautuneisuutta, jolloin niillä saadut tulokset eivät välttämättä anna luotettavinta tulosta, jos muuttuja ei olekaan normaalijakautunut.

Testin lähtökohtana on laskea otoksen sekä normaalijakauman *kumulatiivisten suhteellisten frekvenssijakaumien* suurinta erotusta. Toisin sanoen verrataan kumulatiivisia suhteellisia frekvenssejä toisiinsa ja etsitään maksimipoikkeama niiden väliltä. Tämä voidaan esittää kaavalla

$$D = \max |F_0(x_i) - S_N(x_i)|, \quad \text{i} = 1,\, 2,\, ...,\, \text{N},$$

missä D on maksimipoikkeama, F_0 on normaalijakauman kumulatiivinen suhteellinen frekvenssi, S_N on otoksen kumulatiivinen suhteellinen frekvenssi ja N on otoksen koko. Saatu maksimipoikkeama D on testisuure, jonka avulla määritetään otoksen sopivuus jakaumaan. Kumulatiivinen frekvenssi saadaan kerryttämällä eri muuttujien frekvenssejä pienimmästä suurimpaan. Toisin sanoen kunkin muuttujan arvon kohdalla lasketaan kuinka monta tilastoyksikköä on saanut pienemmän tai yhtäsuuren arvon. Tarkastellaan tätä esimerkkiaineiston avulla. Taulukossa 4.2 on esitetty esimerkkiaineiston arvosanojen jakautuminen opiskelijoittain eli kunkin arvosanan frekvenssi. Nyt jokaisen arvosanan kohdalle lasketaan kumulatiivinen frekvenssi ja esitetään se taulukossa 4.4. Taulukkoon on laskettu myös otoksen sekä normaalijakauman kumulatiivinen suhteellinen frekvenssi sekä näiden erotuksen itseisarvo.

Taulukko 4.4: Arvosanojen jakautuminen ja Kolmogorovin–Smirnovin testin suureet.

Arvosana	f	kumul. f	S_N	F_0	$\lvert F_0 - S_N \rvert$
0	1	1	0,091	0,088	0,003
1	1	2	0,182	0,196	0,014
2	2	4	0,364	0,359	0,004
3	4	8	0,727	0,554	0,174
4	2	10	0,909	0,736	0,173
5	1	11	1,000	0,870	0,130

Kumulatiivisesta frekvenssistä saadaan suhteellinen skaalaamalla se välille $[0, 1]$, eli jakamalla kumulatiivisen frekvenssin arvot otoksen koolla. Normaalijakauman kumulatiivinen suhteellinen frekvenssi saadaan joko laskemalla se jakaumasta sopivalla keskiarvolla ja varianssilla. Esimerkissä käytettiin normaalijakaumalle tenttiarvosanan keskiarvoa $\mu = 2,727$ ja varianssia $\sigma^2 = 2,018$. Suurin poikkeaman arvo, eli laskettu testisuure, on $D = 0,174$. Koska testissä ei ole käytetty ennalta määrättyjä keskiarvon ja varianssin arvoja, vaan ne on laskettu otoksesta, kutsutaan normaalijakauman tapauksessa testiä *Kolmogorv-Smirnov-Lillieforsin testiksi* tai lyhyemmin *Lillieforsin testiksi.*

Testisuureen tulkinta käsitellään esimerkkinä seuraavassa aliluvussa, jossa käsitellään tarkemmin testisuureiden testausta. Kolmogorovin–Smirnovin testissä testisuureen tulkinta perustuu taulukkoarvoihin. Lillieforsin testin yhteydessä käytetään Lillieforsin laskemia taulukkoarvoja, jotka on määritetty erityisesti testaamaan aineiston sopivuutta normaalijakaumaan, jonka keskiarvo ja varianssi eivät ole ennalta tiedossa.

4.2.2 Hypoteesin testaus

Vaikka mitatun aineiston muuttujat eivät olisikaan jakautuneet minkään tunnetun jakauman perusteella, niin jakaumia hyödynnetään *hypoteesin testaamisessa*, jota käsitellään seuraavaksi. Hyvin määritellyt tutkimuskysymykset ovat sellaisia, jotka voidaan testata. Tällöin tutkittava asia on muotoiltava hypoteesiksi, joka koostuu kahdesta osasta: *nollahypoteesista* H_0 sekä sen *vastahypoteesista* H_1. Nollahypoteesi määritellään yleensä sellaiseksi väitteeksi, jonka tutkimus pyrkii hylkäämään, eli esimerkiksi “Opiskelijoiden harjoitustehtävien teolla ei ole vaikutusta tenttisuoritukseen.” Vastahypoteesi olisi tällöin “Opiskelijoiden harjoitustehtävien teko vaikuttaa tenttisuoritukseen.”

Tarkoituksena on siis pyrkiä aineistosta ja tutkimuksesta riippuen hylkäämään tai hyväksymään nollahypoteesi. Oletetaan, että harjoitustehtävien teolla ja tenttisuorituksella on haivaittu olevan yhteys, jota ollaan testaamassa. Yhteys voi olla todellista tai se voi olla näennäistä, jolloin se on johtunut niin sanotusta *satunnaisvirheestä*. Satunnaisvirheen voidaan oletaa jakautuneen normaali- tai t-jakauman mukaisesti, jolloin aineistosta voidaan laskea ilmiötä kuvaava *testisuure*. Testisuureen ja jakauman avulla voidaan laskea todennäköisyys sille, että satunnaisvirhe on poikkeuksellisen voimakasta eli havaittu yhteys johtuu muusta kuin satunnaisvirheestä ja näin ollen nollahypoteesi voidaan hylätä. Tätä todennäköisyyttä kutsutaan *p-arvoksi* ja sen avulla saadaan niin sanottu *riskitaso* nollahypoteesin hylkäämiselle. Riskitaso kuvaa todennäköisyyttä, eli riskiä virheelliselle nollahypoteesin hylkäämiselle.

Tutkimuksen yhteydessä p-arvo voidaan ilmoittaa joko suoraan lasketun testisuureen kohdalla tai käyttää kiinteitä p-arvoja. Yleisimmin käytetyt kiinteät p-arvot ja niiden tulkinnat on kuvattu taulukossa 4.5. Sanalliset tulkinnat vastaavat kysymykseen “Onko tutkittavan ilmiön vaikutus tilastollisesti merkitsevä satunnaisvirheeseen nähden?”

Taulukko 4.5: Kiinteiden p-arvojen riskitasot, sanalliset kuvaukset sekä taulukoissa käytetyt merkinnät.

p-arvo	Riskitaso	Sanallinen kuvaus	Taulukkomerkintä
$p < 0.001$	0,1 %	erittäin merkitsevä	***
$p < 0.01$	1 %	merkitsevä	**
$p < 0.05$	5 %	melkein merkitsevä	*

Ihmisten käytöstä ja toimintaa tutkittaessa 5 % riskitaso riittää hyvin, vaikka sanallinen kuvaus “melkein merkitsevä” antaakin hieman harhaanjohtavan kuvan. Tiukempia riskitasoja käytetään tilanteissa, joissa halutaan varmistua, ettei nollahypoteesia hylätä liian vähin perustein. Tarkastellaan esimerkkinä edellisessä aliluvussa käsiteltyä Kolmogorovin–Smirnovin testiä ja sen testisuuretta D. Testisuure sai esimerkissä arvon $D = 0,174$. Tarkoituksena on selvittää poikkeaako otoksen kumulatiivinen suhteellinen jakauma merkitsevästi normaalijakauman vastaavasta jakaumasta vai ei. Testisuuretta, eli maksimipoikkeamaa, verrataan Kolmogorovin–Smirnovin tai Lillieforsin taulukoimiin kriittisiin poikkeaman arvoihin. Kriittisten arvojen taulukot on esitetty liitteessä D.2. Käytettäessä 5 % merkitsevyystasoa, Lillieforsin kriittinen arvo otoskoolla $N = 11$ on 0,249. Koska laskettu testisuure on pienempi kuin kriittinen arvo, voidaan päätellä, että otos ei eroa merkitsevästi normaalijakaumasta. Näin ollen voidaan olettaa esimerkkiaineiston tenttiarvosanojen jakautuvan normaalisti. Toinen esimerkki hypoteesin testaamisesta, testisuureesta ja p-arvosta on esitetty aliluvussa 4.2.4 sivulla 30.

4.2.3 Lineaarinen regressio

Yleisesti tutkimuksissa mitataan yksittäistä tilastoyksikköä useamman kuin yhden muuttujan suhteen, jolloin aineistosta halutaan tutkia muuttujien muuttujien välisiä yhteyksiä. Tarkoituksena on luoda matemaattinen malli muuttujien muodostamasta yhteydestä. Eräs tällainen menetelmä on *regressioanalyysi*, jonka yksinkertaisin malli on *lineaarinen regressio*. Kahden muuttujan y ja x välistä suhdetta kuvaava malli saadaan suoran yhtälöstä

$$y = \alpha + \beta x + \varepsilon,$$

missä α on y-akselin leikkauspiste, β on suoran kulmakerroin ja ϵ on satunnaisvirhe, jonka voidaan siis olettaa olevan satunnaismuuttuja ja jakautunut normaalijakauman mukaisesti. Jos tutkitaan ilmiötä, josta ei ole saatavilla mallille valmiita parametrejä α ja β, niin ne tulee määrittää. Tällöin pyritään sovittamaan mitattuun aineistoon mahdollisimman hyvin sopiva suora. Suoraa kutsutaan *sovitetuksi malliksi* ja se tulee erottaa teoreettisesta "oikeasta" mallista. Tämän vuoksi mallin parametrit muutetaan muotoon $\hat{y}_i = a + bx_i$. Sovittamiseen käytetään *pienimmän neliösumman menetelmää*, johon tutustutaan seuraavaksi

Aluksi määritellään sovitetun mallin virhe e_i eli niin sanottu *residuaali* eli *jäännöstermi*.

Määritelmä 4.2.3. Kun tiedetään aineisto $\{(x_i, y_i) \,|\, i = 1, 2, ...n\}$ ja sovitettava malli $\hat{y}_i = a + bx_i$, niin i. jäännöstermi e_i on

$$e_i = y_i - \hat{y}_i.$$

Määritelmän avulla saadaan muuttujien välille hyödyllinen yhteys:

$$y_i = a + bx_i + e_i.$$

Tarkoituksena on löytää parametrit a ja b siten, että jäännöstermien neliöiden summa olisi mahdollisimman pieni. Tästä käytetään myös nimitystä virheen neliöiden summa (the sum of squares of the errors) ja kyseistä suuretta merkitään lyhenteellä *SSE*. Toisin sanoen pienimmän neliösumman menetelmää käytetään siis parametrien

a ja b etsimiseksi minimoimalla summaa

$$SSE = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = \sum_{i=1}^{n} (y_i - a - bx_i)^2.$$

Osittaisderivoimalla summaa SSE parametrien a ja b suhteen saadaan

$$\frac{\partial(SSE)}{\partial a} = -2\sum_{i=1}^{n}(y_i - a - bx_i) \quad \text{ja} \quad \frac{\partial(SSE)}{\partial b} = -2\sum_{i=1}^{n}(y_i - a - bx_i)x_i.$$

Asettamalla osittaisdifferentiaalit nolliksi saadaan minimikohtia kuvaavat yhtälöt

$$na + b\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i \quad \text{ja} \quad a\sum_{i=1}^{n} x_i + b\sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} x_i y_i,$$

joista voidaan määrittää parametreille a ja b laskukaavat.

Määritelmä 4.2.4. Kun tiedetään aineisto $\{(x_i, y_i) \mid i = 1, 2, ...n\}$, niin pienimmän neliösumman mukaiset arvot a ja b regressioparametreille α ja β saadaan kaavoista

$$b = \frac{n\sum_{i=1}^{n} x_i y_i - \left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} y_i\right)}{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

ja

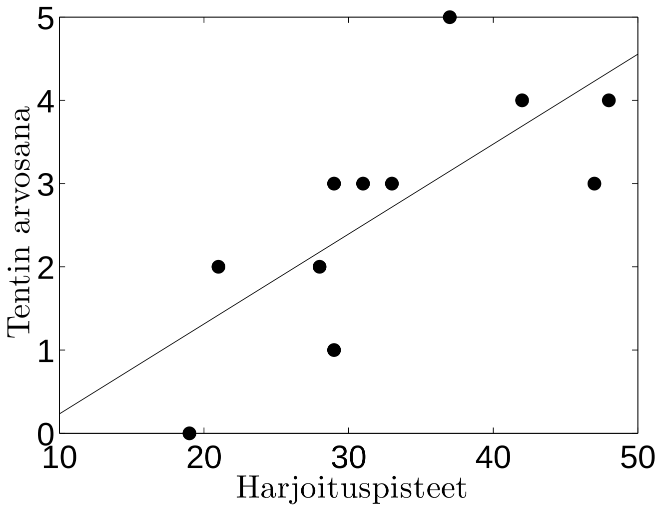
$$a = \frac{\sum_{i=1}^{n} y_i - b\sum_{i=1}^{n} x_i}{n} = \bar{y} - b\bar{x},$$

missä $\bar{x}$ ja $\bar{y}$ ovat muuttujien x ja y keskiarvot.

Näin saatua regressiosuoraa kuvataan yhdessä muuttujien *sirontakuvion* kanssa. Otetaan esimerkiksi tutun esimerkkiaineiston (taulukko 4.1) harjoituspisteiden ja tentin arvosanojen välinen sirontakuvio sekä näille muuttujille laskettu regressiosuora. Sirontakuvio ja regressiosuora on esitetty kuvassa 4.5.

Regressioanalyysi kattaa paljon muitakin kuin lineaarisia malleja, joita voidaan hyödyntää erilaisille aineistoille. Lineaarinen regressio on kuitenkin hyvä lähtökohta muuttujien väliselle tutkimukselle ja varsinkin sirontakuviosta voidaan havaita tarvittavan sovitemallin muotoja.

Lineaarisen regression avulla voidaan myös määrittää muuttujien *selitysaste* eli luku, joka kertoo kuinka paljon esimerkiksi muuttujasta y on selitettävissä muuttujalla x. Tarkalleen ottaen kyse on siitä kuinka suuri osa varianssista selittyy sovitetulla mallilla. Selitysasteen laskemiseen tarvitaan aikaisemmin määritellyn vir-



Kuva 4.5: Esimerkkiaineiston sirontakuvio harjoituspisteiden ja tentin arvosanojen suhteen sekä lineaarinen regression sovitesuora.

heen neliöllisen summan SSE lisäksi myös *todellista neliöiden summaa* (total sum of squares), josta käytetään merkitää SST, joka lasketaan kaavalla

$$SST = \sum_{i=1}^{n} (y_i - \bar{y})^2.$$

Selitysaste R^2 lasketaan kaavalla

$$R^2 = 1 - \frac{SSE}{SST} = 1 - \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2}$$

Jos virhettä ei esiinny, niin $SSE = 0$ ja tällöin selitysaste on $R^2 = 1$. Toisaalta, jos SSE on vain vähän pienempi kuin SST, niin selitysaste $R^2 \approx 0$. Esimerkkiaineiston tapauksessa selitysasteeksi saadaan $R^2 = 0.5314$, jolloin harjoitustehtävien tekeminen selittäisi noin puolet tenttiarvosanasta. Selitysasteen tilastollisesta merkitsevyydestä tarkemmin seuraavan luvun esimerkissä.

4.2.4 Korrelaatio

Toinen yksinkertainen tapa saada suuntaa antavia tietoja muuttujien välisistä yhteyksistä on *korrelaatio*. Seuraavaksi esitellään kaksi erilaista menetelmää määrittää kahden muuttujan välinen korrelaatio. Ensimmäiseksi käsitellään *Pearsonin tulomomenttikorrelaatiokerroin* (Pearson product-moment correlation coefficient), joka on niin sanottu *parametrillinen* menetelmä eli se olettaa muuttujien noudattavan normaalijakaumaa. Tätä ei kuitenkaan voida aina olettaa, joten esitellään myös parametrittomista menetelmistä *Spearmanin järjestyskorrelaatiokerroin* (Spearman's rank correlation coefficient).

Pearsonin tulomomenttikorrelaatiokerroin

Pearsonin tulomomenttikorrelaatiokerrointa laskiessa muuttujien tulee olla mitattu vähintään välimatka-asteikolla. Ennen korrelaation laskemista määritellään sen laskemiseen tarvittava *kovarianssi*, joka kuvaa kahden muuttujan yhteyttä.

Määritelmä 4.2.5. Muuttujien x_1 ja x_2 välinen kovarianssi s_{12}, lasketaan kaavalla

$$s_{12} = \frac{1}{n-1}\sum_{j=1}^{n}(x_{1j} - \bar{x}_1)(x_{2j} - \bar{x}_2),$$

missä x_{ij} on muuttujan x_i j:s havaintoarvo ($j = 1, 2, 3, ..., n$), $\bar{x}_i$ on muuttujan x_i keskiarvo ja n on otoksen koko.

Aikaisemmin määritellyllä varianssilla ja kovarianssilla on yhteys, sillä laskemalla kovarianssi muuttujan itsensä kanssa, saadaan muuttujan varianssi. Kovarianssi kuvaa muuttujan havaintoarvojen keskinäistä suhdetta keskiarvoon. Jos molemmat havaintoarvot poikkeavat samaan suuntaan muuttujansa keskiarvosta, niin erotusten tulo on positiivinen ja vastaavasti havaintoarvojen ollessa eri puolilla keskiarvoa, summan termi on negatiivinen. Summaamalla nämä termit saadaan muuttujia kuvaava yhteys. Skaalaamalla kovarianssi välille $[-1, 1]$ saadaan haluttu muuttujien välinen korrelaatiokerroin. Tämä tapahtuu jakamalla kovarianssi muuttujien hajontojen tulolla.

Määritelmä 4.2.6. Muuttujien $x_1$ja x_2 välinen (Pearsonin) korrelaatio r_{12} lasketaan kaavalla

$$r_{12} = \frac{s_{12}}{\sqrt{s_{11}}\sqrt{s_{22}}} = \frac{s_{12}}{\sqrt{s_1^2}\sqrt{s_2^2}} = \frac{s_{12}}{s_1 s_2}.$$

Korrelaatiokerroin saa siis arvoja välillä $[-1, 1]$ ja lineaarinen yhteys muuttujien välillä on sitä suurempi, mitä suurempi korrelaatiokertoimen itseisarvo on. Positiivinen korrelaatio tarkoittaa sitä, että esimerkiksi muuttujan x_1 arvon kasvaessa kasvaa myös muuttujan x_2 arvo ja päinvastoin. Negatiivinen korrelaatio tarkoittaa vastaavasti toisen muuttujan arvon pienenemistä toisen kasvaessa. Korrelaatiokertoimen r_{12} neliö r_{12}^2 tarkoittaa samaa selitysosuutta kuin edellisessä luvussa lineaarisen regression yhteydessä määritelty selitysosuus R^2. Selitysosuus kertoo siis kuinka paljon muuttujan varianssista on selitettävissä toisen muuttujan avulla. Taulukon 4.1 esimerkkiaineistolle korrelaatiokerroin harjoituspisteiden ja tentin arvosanan perusteella olisi $r_{12} = 0,7290$ ja selitysosuus $r_{12}^2 = 0.5314$, joka tarkoittaa, että kyseisen esimerkkiaineiston kohdalla ahkera harjoitustehtävien tekeminen on tuottanut parempia tuloksia tentissä. Noin puolet tenttiarvosanasta on selitettävissä harjoitustehtävien tekemisellä.

Korrelaatiokerroin kertoo ainoastaan muuttujien laskennallisesta yhteydestä, eikä yhteyden tulkitseminen ainoastaan korrelaatiokertoimen avulla ei ole järkevää. Pelkän korrelaatiokertoimen avulla ei voida päätellä kumpi muuttujista mahdollisesti on syynä toiseen muuttujaan. Joidenkin muuttujien kohdalla vaikutusta voidaan tutkia tarkastelemalla tapahtumia ajallisesti, jolloin ensin tapahtunut vaikuttaa jälkimmäiseen. Esimerkiksi harjoitustehtäviä tehdään ennen tenttiä, jolloin harjoitustehtävien tekeminen vaikuttaa tenttiarvosanaan, eikä päinvastoin. Lisäksi otoksen koko vaikuttaa siihen, voidaanko korrelaatiokertoimen suuruudesta päätellä merkittävää yhteyttä muuttujien välille. Korrelaatiokertoimen tilastollisesta merkitsevyyttä voidaan tutkia t-jakauman avulla laskemalla korrelaatiokertoimen r ja vapausasteiden $n-2$ mukainen testisuure t kaavalla

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}},$$

missä n on otoksen koko. Esimerkkiaineiston kohdalla vapausastein $11-2=9$ saadaan $t = 3,1947$. Nollahypoteesina on nyt “Korrelaatiokerroin ei eroa nollasta.” Laskemalla t-arvoa vastaava p-arvo t-jakaumasta vapausastein $n-2$ saadaan $p = 0.0109$. Tulkitsemalla p-arvoa taulukon 4.5 mukaisesti voidaan nollahypoteesi hylätä 5 %:n riskitasolla, jolloin korrelaatio eroaa nollasta melkein merkitsevästi. Tällaisessa tilanteessa on järkevää merkitä p-arvo näkyviin ja jättää sen tulkinta lukijalle, koska se on niin lähellä 1 %-riskitasoa.

Spearmanin järjestyskorrelaatiokerroin

Spearmanin korrelaatiokerrointa laskiessa muuttujien tulee olla mitattu vähintään järjestysasteikolla. Korrelaation laskemiseen käytetään määritelmän 4.2.6 mukaista laskukaavaa. Spearmanin korrelaatiokertoimesta käytetään merkintää ρ. Erona Pearsonin menetelmään on se, että muuttujan alkuperäisten arvojen sijaan käytetäänkin järjestysnumeroita. Järjestysnumerot saadaan järjestämällä muuttuja suuruusjärjestykseen.

Järjestämällä taulukon 4.1 esimerkkiaineisto suuruusjärjestykseen sekä harjoituspisteiden että tenttiarvosanan perusteella saadaan muuttujille järjestysnumerot, jotka on esitetty taulukossa 4.6.

Laskemalla järjestysnumeroiden avulla määritelmän 4.2.6 mukainen korrelaatio saadaan $\rho_{12} = 0,8270$. Tämä on suurempi kuin Pearsonin korrelaatiokertoimen arvo. Korrelaatiokertoimen tilastollisesta merkitsevyyttä voidaan tutkia t-jakauman avulla, kuten tehtiin Pearsonin korrelaationkin tapauksessa. Tällöin saadaan vapausastein 9 t-arvoksi $t = 4,413$, joka vastaa p-arvoa $p = 0.0017$. Tällöin voidaan

Taulukko 4.6: Esimerkkidata lisättynä harjoituspisteiden sekä tenttiarvosanan järjestysnumeroilla.

Opiskelija-numero	Koulutus-ohjelma	Vuosi-kurssi	Harjoitus-pisteet	Tentin arvosana
9551	Sähkötekniikka	2	37(8)	5(11)
3080	Konetekniikka	2	33(7)	3(6,5)
6461	Biotekniikka	6	48(11)	4(9,5)
5373	Tietotekniikka	3	31(6)	3(6,5)
9021	Tuotantotalous	2	21(2)	2(3,5)
7858	Konetekniikka	3	29(4,5)	3(6,5)
5108	Sähkötekniikka	4	42(9)	4(9,5)
1166	Tietotekniikka	5	28(3)	2(3,5)
8392	Konetekniikka	2	29(4,5)	1(2)
5002	Tietotekniikka	4	47(10)	3(6,5)
6749	Tuotantotalous	3	19(1)	0(1)

todeta korrelaation olevan tilastollisesti merkitsevä.

4.2.5 Klusterianalyysi

Aineiston kuvailu ja muuttujien välisten suhteiden tarkastelu tutkimuksen alkuvaiheessa antaa hyvän pohjan muihin hyödyllisiin tilastollisiin menetelmiin. Usein tutkimuksessa halutaan ryhmitellä samankaltaiset tilastoyksiköt samaan ryhmään ja toisaalta erotella näistä poikkeavat havainnot omiin ryhmiinsä. Tässä tutkimuksessa opiskelijoita haluttiin ryhmitellä jumppakäytöksen ja sitä seuranneen oppimisen perusteella ryhmiin. Ryhmittelyn menetelmäksi valittiin klusterianalyysi, jota esitellään seuraavaksi lyhyesti.

Klusterianalyysi tai ryhmittelyanalyysi jakaa havaintoyksiköt ryhmiin haluttujen muuttujien perusteella. Etukäteen ei tarvitse tietää ryhmittelyperustetta, vaan riittää määrittää muuttujat, joiden suhteen ryhmittely halutaan tehdä. Klusterointi on siis aineiston tutkimisen, eikä niinkään teorian vahvistamisen työkalu. Klusterianalyysissa on paljon erilaisia menetelmiä ja algoritmeja ryhmittelyn tekemiseen ja niiden valinta perustuu pitkälti käytettävissä olevasta aineistosta ja halutusta lopputuloksesta, kuten klustereiden määrästä ja suhteellisesta koosta. Kaksi yleisimmin käytettyä klusterointimenetelmää ovat *hierarkinen klusterointi* (hierarchial clustering) ja *K-means klusterointi* (K-means clustering), jota esimerkiksi Metsämuurosen [13] suomentamana voidaan kutsua myös K-keskiarvon klusteroinniksi. Näiden lisäksi on myös muun muassa jakaumapohjainen klusterointi [18] (distribution-based clustering) ja tiheyspohjainen klusterointi [9](density-based clustering) ja paljon muita menetelmiä, joihin ei kuitenkaan perehdytä tässä tutkielmassa.

Yleisesti ryhmittelyanalyysi voidaan jakaa kolmeen vaiheeseen. Ensimmäisessä vaiheessa muuttujat on standardoitava tai muulla tavalla skaalattava, jotta suuren vaihteluvälin muuttujat eivät vaikuttaisi pienen vaihteluvälin muuttujia enempää klusteroinnin tuloksiin. Tämä valinta on tehtävä muuttujakohtaisesti käytetyn mitta-asteikon ja muuttujan luonteen mukaisesti. Standardointi tehdään satunnaismuuttujalle vähentämällä kustakin muuttujan havaintoarvosta odotusarvo ja jakamalla erotus keskihajonnalla. Havaintoarvon x *standardoiduksi arvoksi* z saadaan siis

$$z = \frac{x - \mu}{\sigma}.$$

Tämän lisäksi voidaan käyttää erilaisia, tilanteeseen sopivia tapoja skaalata muuttujan havaintoarvot esimerkiksi välille $[0, 1]$. Tämän voi toteuttaa esimerkiksi jakamalla muuttujan havaintoarvot vaihteluvälin tai mitta-asteikon suurimmalla arvolla.

Toinen vaihe on kahden ryhmiteltävän alkion etäisyyden mittaaminen. Etäisyyden mittaamiseen on kiinnitettävä huomiota, koska se kuvaa alkioiden läheisyyttä tai samanlaisuutta. Erilaisilla muuttujilla etäisyys voidaan laskea eri tavoin tapauksesta riippuen. Helpoin ja ymmärrettävin tapa etäisyyden laskemiseen on *Euklidinen etäisyys*, joka on siis kahden alkion välille piirretyn janan pituus. Jos käytössä on p kappaletta muuttujia, niin alkioiden x ja y välinen Euklidinen etäisyys $d(x, y)$ lasketaan kaavalla

$$d(x, y) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \cdots + (x_p - y_p)^2},$$

missä x_i ja y_i ovat muuttujan i mukaiset havaintoarvot tilastoyksiköille. Muuttujat ovat siis enemmän samanlaiset, jos niiden erot eri muuttujien suhteen ovat mahdollisimman pieniä. Tämän lisäksi käytetään yleisesti edellisen neliötä, jolloin kaava saadaan vielä yksinkertaisempaan muotoon

$$d(x, y) = (x_1 - y_1)^2 + (x_2 - y_2)^2 + \cdots + (x_p - y_p)^2,$$

joka on helpompi havainnollistaa ja laskea, kun ei tarvitse suorittaa laskuteknisesti hidasta neliöjuurioperaatiota. Yleistetty versio Euklidisesta etäisyydestä on *Minkowskin metriikka*

$$d(x, y) = \left[\sum_{i=1}^{p} |x_i - y_i|^m \right]^{1/m}, \qquad m \geq 1,$$

mistä saadaan Euklideen etäisyys kun $m = 2$. Minkowskin metriikalla voidaan parametrin m avulla painottaa suurien ja pienien etäisyyksien painoa.

Kolmas vaihe klusteroinnissa on klustereiden välisen etäisyyden laskeminen. Tähän käytetään erilaisia menetelmiä klusterointitavasta riippuen. Tarkasteltaessa tarkemmin hierarkista klusterointia esitellään siihen liittyvät menetelmät. K-means klusteroinnissa puolestaan klusterin sijainti on niin sanotussa klusterikeskuksessa, johon etäisyys voidaan laskea käyttämällä toisessa vaiheessa esiteltyjä menetelmiä.

Hierarkinen klusterointi

Hierarkisen klusteroinnin lähtökohtana on, että kaikki alkiot ovat aluksi omia klustereitaan, joita yhdistelemällä saadaan yksi suuri klusteri. Tätä kutsutaan *kokoavaksi hierarkiseksi metodiksi* (agglomerative hierarchical method). Toinen mahdollisuus, eli *erilaistava hierarkinen metodi* (Divisive hierarchical method) lähtee oletuksesta, että alkiot muodostavat aluksi yhden klusterin, jota sitten lähdetään jakamaan. Esitellään seuraavaksi tarkemmin kasaavan metodin algoritmi, kun otoskoko on N alkiota.

1. Aluksi klustereita on N kappaletta, joista jokainen on oma itsenäinen kokonaisuutensa. Näille lasketaan symmetrinen $N \times N$ etäisyysmatriisi $\mathbf{D} = \{d_{ik}\}$.

2. Etsitään etäisyysmatriisista toisiaan lähinnä oleva klusteripari. Olkoon kahden lähimmän klusterin U ja V välinen etäisyys d_{UV}.

3. Yhdistetään klusterit U ja V uudeksi klusteriksi (UV). Päivitetään etäisyysmatriisin alkiot poistamalla ensin klustereiden U ja V rivit ja sarakkeet etäisyysmatriisista ja lisätään tämän jälkeen uuden klusterin (UV) etäisyyksiä vastaavat rivi ja sarake etäisyysmatriisiin.

4. Toistetaan kohtia 2 ja 3 $N-1$ kertaa, jolloin kaikki alkiot ovat yhdessä klusterissa. Kirjataan yhdistetyt klusterit ja etäisyystasot, joilla yhdistäminen tehtiin.

Analyysia varten haluttu määrä klustereita saadaan "peruuttamalla" algoritmissa tarvittava määrä, jolloin viimeisiä yhdistämisiä ei tehdä. Tällä algoritmilla ei kuitenkaan ole takeita klustereiden koon suhteen vaan poikkeavat alkiot saatetaan yhdistää vasta viimeisillä algoritmin kierroksilla. Tämä voi toki olla myös haluttu ilmiö esimerkiksi poikkeavia alkioita etsittäessä. Yhdistettyjen klustereiden väliseen etäisyyteen on hierarkisessa klusteroinnissa olemassa kolme erilaista tapaa: *lähimmän naapurin menetelmä*, *kaukaisimman naapurin menetelmä* sekä *naapurikeskiarvon menetelmä*. Klustereiden välinen etäisyys lähimmän naapurin menetelmässä on lyhin etäisyys, joka klustereiden alkioilla on. Olkoon kaksi klusteria (UV) ja W, jolloin saadaan

$$d_{(UV)W} = \min\{d_{UW}, d_{VW}\},$$

missä valitaan siis lyhin etäisyyksistä d_{UW} ja d_{VW}. Tästä laskentaa voidaan rekursiivisesti jatkaa, jos kyse on useamman alkion klustereista. Tämän perusteella on helppoa ymmärtää kaukaisimman naapurin menetelmän tapa laskea klustereiden välinen etäisyys. Olkoon klusterit (UV) ja W kuten edellä, jolloin

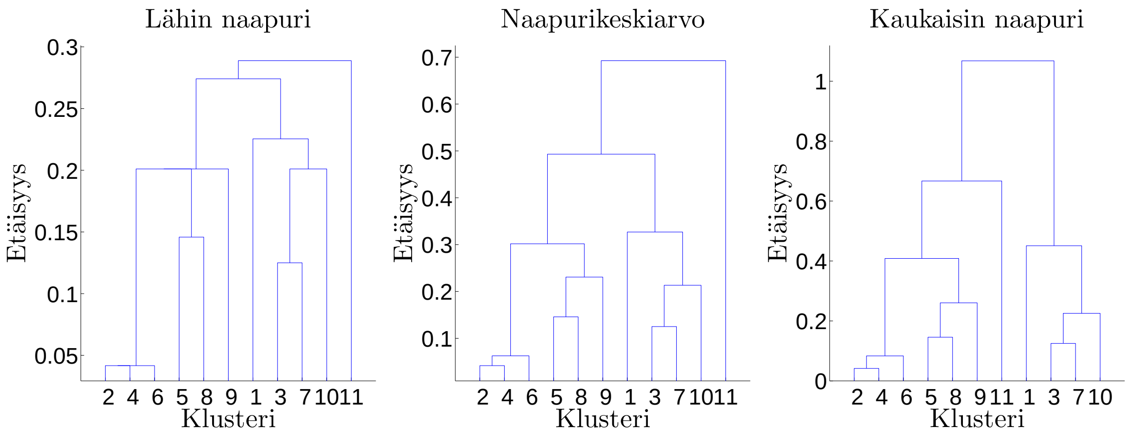
$$d_{(UV)W} = \max\{d_{UW}, d_{VW}\}.$$

Kolmas tapa laskea etäisyys, eli naapurikeskiarvon menetelmä, laskee klustereiden välisten alkioiden etäisyyksien keskiarvon ja jakaa sen etäisyyksien määrällä. Olkoon klusterit (UV) ja W jälleen kuten edellä ja lisäksi klusterissa (UV) on $N_{(UV)}$ alkiota ja klusterissa W on N_W alkiota. Tällöin klustereiden väliseksi etäisyydeksi saadaan

$$d_{(UV)W} = \frac{\sum_i \sum_j d_{ik}}{N_{(UV)} N_W},$$

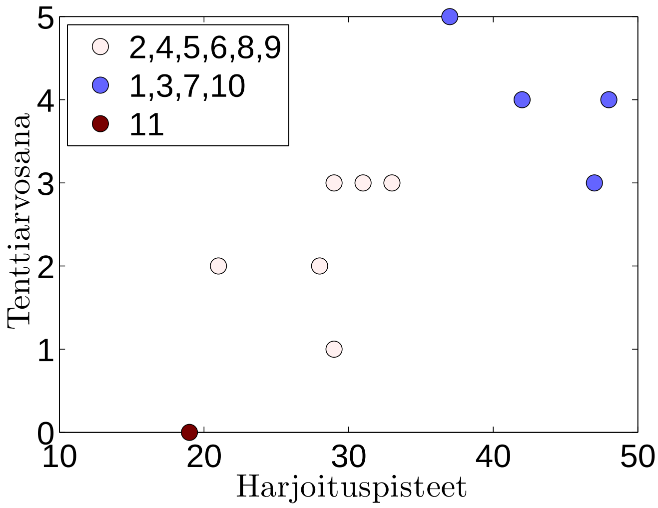
missä $i = 1, ..., N_{(UV)}$ ja $k = 1, ..., N_W$. Näiden erilaisten tapojen avulla saadaan erilaisia tuloksia, jotka sopivat tietynlaisille aineistoille. Kaukaisimman naapurin menetelmä tuottaa tiukkarajaisempia ja kompaktimpia klustereita kuin lähimmän naapurin menetelmä. Lähimmän naapurin menetelmä on kuitenkin monipuolisempi, ja se voi erottaa esimerkiksi sisäkkäisiä klustereita toisistaan.

Hierarkista klusterointia visualisoidaan käyttämällä *dendrogrammeja*. Kuvassa 4.6 on esitetty taulukon 4.1 esimerkkiaineiston hierarkisen klusteroinnin dendrogrammit.



Kuva 4.6: Esimerkkiaineiston hierarkisen klusteroinnin dendrogrammit käyttäen lähimmän ja kaukaisimman naapurin sekä naapurikeskiarvon menetelmiä.

Kuvassa klusterointi on tehty käyttämällä kolmea edellä esitettyä klustereiden välisen etäisyyden laskemisen menetelmää. Klusterit on numeroitu aluksi siinä järjestyksessä kuin alkiot on taulukossa esitetty ja tämän jälkeen etsitty lyhin kahden alkion välinen etäisyys ja yhdistetty kyseiset pisteet. Esimerkiksi kuvasta nähdään, että ensimmäisenä on yhdistetty klusterit kaksi ja neljä, jonka jälkeen tähän klusteriin on lisätty kuudes alkio ja niin edelleen. Jos halutaan jakaa aineisto esimerkiksi kolmeen ryhmään, niin katsotaan kuvaa ylhäältä alaspäin ja vedetään viiva kahden viimeisen yhdistämisen alapuolelle, jolloin saadaan kolme erillistä klusteria. Esimerkkiaineiston tapauksessa kaikilla menetelmillä on saatu samat kolme klusteria. Kuvassa 4.7 on esitetty klusteroinnin tulos piirtämällä alkiot harjoituspisteiden ja tenttiarvosanan mukaisesti.



Kuva 4.7: Klusteripisteet hierarkisella klusteroinnilla.

Kuvasta nähdään, että klusteri 11 muodostaa yksin yhden klusterin. Aineistosta riippuen näistä voidaan havaita aineistosta selvästi poikkeavia tilastoyksiköitä, joita tilanteesta riippuen voidaan joko poistaa tai tulkita aineistoon kuuluvaksi. Klusteroinnin jälkeen klustereita voidaan vertailla, analysoida ja nimetä ominaisuuksiensa perusteella.

K-means klusterointi

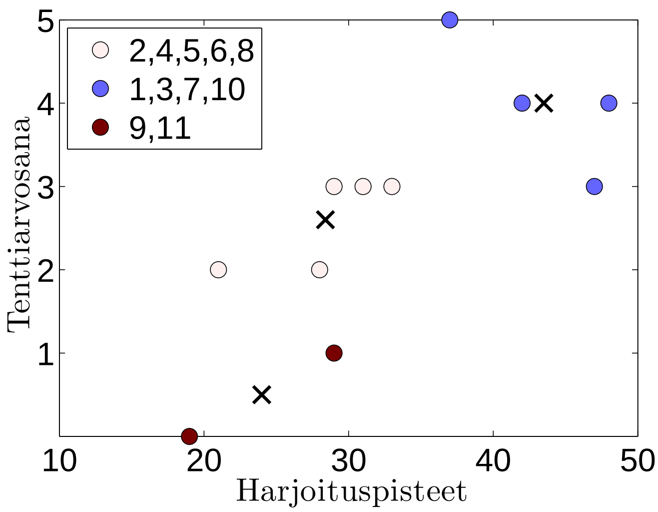
Toinen yleisesti käytetty klusterointimenetelmä on K-means klusterointi. Siinä lähtökohtana on etsiä klusterikeskuksia aineiston muodostamasta joukosta ja liittää alkiot aina lähimpään klusterikeskukseen. Tämän jälkeen klusterikeskukset päivitetään liitettyjen alkioiden perusteella ja yhdistetään alkiot uudelleen. Tarkastellaan seuraavaksi tarkemmin menetelmän algoritmia.

1. Määritellään K kappaletta klusterikeskuksia. Vaihtoehtoisesti voidaan jakaa alkiot K:hon klusteriin ja laskea näiden perusteella klusterikeskukset kullekkin klusterille.

2. Käydään kaikki alkiot läpi, liitetään ne lähimpään klusterikeskukseen ja lasketaan tämän jälkeen uudelleen klusterikeskukset siihen kuuluvien pisteiden keskiarvona.

3. Toistetaan kohtaa kaksi kunnes alkiot eivät enää vaihda klusterikeskusta.

Klusterikeskuksien sijaintia siis päivitetään siihen kuuluvien pisteiden perusteella, ja uudelleenlaskennan jälkeen alkiot saattavatkin olla lähempänä jotakin toista klusteria kuin ennen uudelleenlaskentaa. Klusterikeskusten alkuarvoja tulee kokeilla useita erilaisia riittävän hyvän klusteroinnin saavuttamiseksi, koska alkuarvoista riippuen voidaan saada erilaisia lopputuloksia. Klustereiden määrä on myös tärkeä aineisto- ja tapauskohtainen tekijä. Klusterointi kannattaa suorittaa myös useammalla K:n arvolla, jotta voidaan löytää paras mahdollinen klusterointi tutkittalle aineistolle.



Kuva 4.8: Klusteripisteet k-means klusteroinnilla sekä klustereiden keskipisteet.

Kuvassa 4.8 on esitetty taulukon 4.1 esimerkkiaineistolle K-means klusterointi arvolla $K = 3$. Kuvassa on esitetty myös lopulliset klusterikeskukset. Jos verrataan hierarkiseen klusterointiin kuvassa 4.7, niin klusterointi tuottaa hieman erilaisen tuloksen. Nyt yhdeksäs alkio on siirtynyt eri klusteriin, jollon 11. alkio ei muodostakaan enää yksin omaa klusteria. K-means klusterointi on muihin klusterointimenetelmiin verrattuna nopea myös suurille aineistoille, vaikka se ei välttämättä laadullisesti annakaan parasta lopputulosta.

5. TULOKSET JA ANALYSOINTI

Tässä luvussa kuvataan tutkimuksessa käytetyt muuttujat ja niiden pohjalta suoritetut analyysit. Luvun ensimmäisessä osassa käsitellään opiskelijakohtaisista muuttujista sekä luokittelu- että suhdemuuttujat. Muuttujista esitellään mitta-asteikot sekä tilastolliset perusmuuttujat. Suhteellisilla asteikoilla mitattavia muuttujia verrataan lisäksi opiskelijaprofiileihin ja lasketaan muuttujien välille korrelaatiokertoimia.

Toisessa osassa käsitellään tehtäviä lokitietojen perusteella ja pyritään havaitsemaan poikkeuksellisen paljon klikkauksia aiheuttaneet tehtävät sekä löytämään syitä klikkausten määrään. Opiskelijoiden ja tehtävien tarkastelu on myös yleiskatsaus tutkimusaineistosta koottuun datamatriisiin.

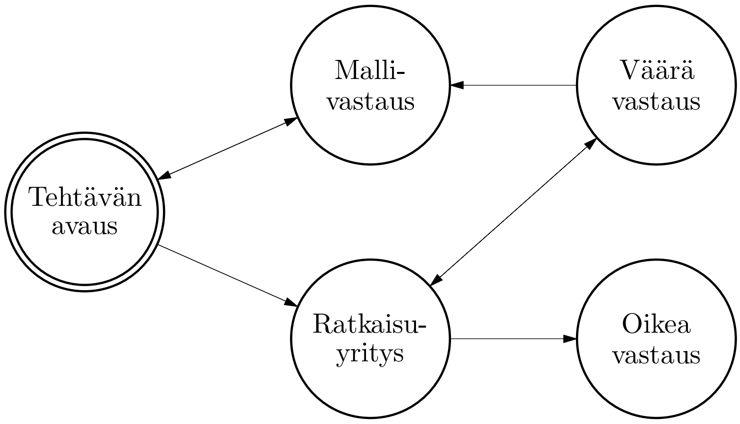
Kolmannessa osassa opiskelijat ryhmitellään klusteroinnin avulla ja tutkitaan klustereiden ominaisuuksia. Tutkimuksesta saatuja havaintoja pyritään vielä lopuksi vahvistamaan muodostamalla ehdollisen todennäköisyyden taulukko ja tutkimalla opiskelijoiden jakautumista muuttujien suhteen.

5.1 Opiskelijat

Tutkimuksessa käytetyt muuttujat on saatu suoraan joko lokitiedoista tai muista opiskelijakohtaisista tiedoista. Lokitiedoista saatavista muuttujista haluttiin valita sellaiset, jotka kuvaisivat mahdollisimman hyvin opiskelijan toimintaa matematiikkajumpassa. Toiminnan määrää pyrittiin aluksi kuvaamaan tehtävän ratkaisuun käytetyn ajan suhteen, mutta lokitiedot eivät mahdollistaneet tarkkaa ja vertailtavissa olevaa aikamuuttujaa. Vaikka tehtävän avaamisen ja oikean tai väärän vastauksen antamisen välinen aika saatiin jopa millisekunnin tarkkuudella, ei mallivastauksen katsomiseen päätyneen opiskelijan käyttämästä ajasta saatu tietoa. Lisäksi Math-Bridge-järjestelmän jäätyä puoleksi tunniksi käyttämättä, aiheutui automaattinen uloskirjautuminen järjestelmästä. Tämä aiheuttaa niin suurta epävarmuutta, ettei tehtäviin käytetyn ajan tai järjestelmässä käytetyn ajan käyttäminen muuttujana ollut järkevää.

Opiskelijakohtaisista tiedoista saatiin neljä luokkamuuttujaa: koulutusohjelma, sukupuoli, opiskelijaprofiili ennen matematiikkajumppaa sekä opiskelijaprofiili matematiikkajumpan jälkeen. Näiden lisäksi saatiin asteikolliset muuttujat: perustaitotestin, matematiikkajumpan jälkeisen jälkitestin tulos sekä Insinöörimatematiikka 1u:n ensimmäisen tentin tulos. Tutkimuksen aikana perustaitotestin ja jälkitestin välinen parannus koettiin paremmaksi muuttujaksi kuin yksittäisten testien tulos. Parannus saatiin vähentämällä jälkitestin tuloksesta perustaitotestin tulos.

Lokitiedoista saatiin opiskelijoille muuttujiksi opiskelijan avaamien tehtävien lukumäärä, tehtävän ratkaisuyritysten lukumäärä, oikeiden vastausten lukumäärä, väärien vastausten lukumäärä, mallivastausten katsomisen lukumäärä sekä kokonaisklikkausten lukumäärä, eli summa avaus-, yritys- sekä malliratkaisun katsomiskerroista. Aina kun opiskelija katsoi mallivastausta, hänen tuli avata tehtävä uudelleen, joka näkyi lokitiedoissa uutena tehtävän avaamisena. Jotta muuttujat saatiin paremmin toisistaan riippumattomiksi, avauskerroista vähennettiin mallivastauksen katsomisen aiheuttamat avauskerrat. Kuvassa 5.1 on esitetty graafi opiskelijan mahdollisista toiminnoista tehtävän avaamisen jälkeen.



Kuva 5.1: Opiskelijan ratkaisupolku tehtävää suoritettaessa.

Opiskelijaa kuvaavien muuttujien lisäksi lokitiedoista saatiin myös tehtäväkohtaisia muuttujia. Opiskelijakohtaisesti määritettiin kunkin tehtävän ensimmäisen avauskerran sekä ensimmäisen oikean ratkaisun aikaleima. Näiden avulla saatiin tutkittua opiskelijoiden tehtävien ratkaisujärjestystä, jota ei kuitenkaan tässä tutkielmassa esitellä. Tämän lisäksi jokaiselle tehtävälle saatiin kokonaisklikkausmäärä, jonka avulla myöhemmin analysoidaan tehtävien aiheuttamaa toimintaa matematiikkajumpassa.

Muuttujien jakaumaa verrattiin normaalijakaumaan laskemalla muuttujille Kolmogorovin–Smirnovin testin mukaiset testisuureet. Testisuureen laskennassa käytettiin muuttujien keskiarvon ja varianssin mukaista normaalijakaumaa, jolloin testi on käytännössä Lillieforsin testi. Lillieforsin ja Kolmogorovin–Smirnovin testin kriittiset arvot on esitelty liittessä D. Taulukossa 5.1 on esitelty kriittiset arvot otoskoolla $N = 118$ ja eri merkitsevyystasoilla.

Taulukosta nähdään, että testisuureen tulkinta riippuu vahvasti valitusta kriittis-

Taulukko 5.1: Lillieforsin ja Kolmogorovin–Smirnovin testin kriittiset testisuureen D arvot otoskoolla $N = 118$ kullakin merkitsevyystasoilla.

Merkitsevyystaso	0,20	0,15	0,10	0,05	0,01
Lilliefors	0,068	0,072	0,074	0,082	0,095
Kolmogorov–Smirnov	0,099	0,105	0,112	0,125	0,150

ten arvojen taulukosta. Lillieforsin taulukko on suunniteltu nimenomaan tutkimuksemme tapaukseen, missä muuttujan havaintoarvoja verrataan ennalta tuntemattomaan normaalijakaumaan. Tällöin varsinaiset johtopäätökset tehdään Lillieforsin kriittisten arvojen perusteella. Tarkastellaan kuitenkin rinnalla myös Kolmogorovin–Smirnovin testin kriittisiä arvoja. Taulukossa 5.2 on esitetty testisuureen arvot tutkimuksessa käytetyille muuttujille, joita mitataan muulla kuin luokitteluasteikolla.

Taulukko 5.2: Tutkimuksessa käytetettyjen muuttujien Lillieforsin testin mukaiset testisuureen arvot.

Muuttuja	Testisuure D
Perustaitotesti	0,233
Jälkitesti	0,117
Parannus	0,082
IMa1u	0,273
Klikkaukset	0,123
Avaukset	0,185
Yritykset	0,134
Mallivastaukset	0,138
Oikeat vastaukset	0,312
Väärät vastaukset	0,145

Verrataan taulukon arvoja ensin Lillieforsin testin kriittiseen arvoon 5 %:n merkitsevyystasolla. Nyt tutkitaan siis vastausta kysymykseen “Poikkeaako aineisto tilastollisesti merkitsevästi normaalijakaumasta?” Testisuureen arvon on oltava pienempi tai yhtä suuri kuin kriittinen arvo. Perustaitotestien välistä parannusta kuvaavan muuttujan testisuure $D = 0,082$, joka on sama kuin Lillieforsin kriittinen arvo. Tällöin se on ainoa muuttuja, joka Lillieforsin testin mukaan normaalijakautunut. Kolmogorovin–Smirnovin määrittelemän kriittisen arvon, $D = 0,125$, perusteella saadaan kuitenkin enemmän normaalijakautuneita muuttujia. Jälkitestin ja klikkausten kokonaismäärän enimmäispoikkeama normaalijakaumasta ovat parannuksen lisäksi pienempiä kuin edellä mainittu kriittinen arvo. Voidaan kuitenkin todeta, että muuttujat eivät Lillieforsin testin perusteella noudata normaalijakaumaa, joten valitaan tutkimusmentelmiksi parametrittomat menetelmät, jotka eivät oleta muuttujilta normaalijakaumaa.

5.1.1 Luokittelumuuttujat

Tutkimuksessa on käytetty neljää luokittelumuuttujaa, koulutusohjelma, opiskelijaprofiilit ennen ja jälkeen matematiikkajumpan sekä sukupuoli. Seuraavaksi käydään läpi luokittelumuuttujat ja opiskelijoiden jakautuminen muuttujien perusteella. Lisäksi jokaisen luokittelumuuttujan kohdalla on selvennetty erilaisten luokitteluun käytettyjen koodien merkitykset.

Koulutusohjelma

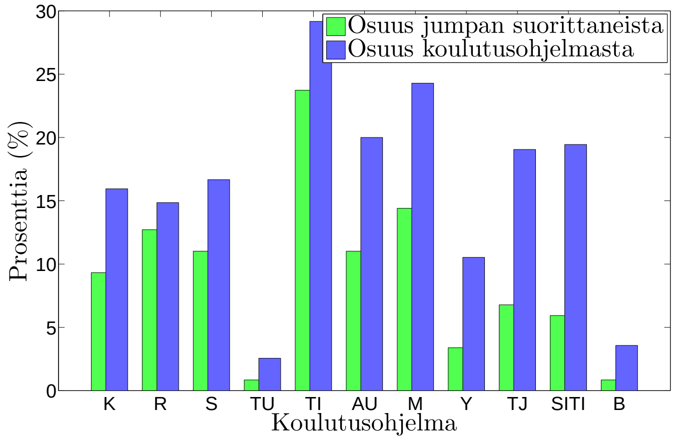
Koulutusohjelma on luokittelumuuttuja, jota mitataan laatueroasteikolla. Aineistossa käytettiin koulutusohjelmista numerokoodeja, jotka vastastaavat koulutusohjelmia seuraavasti:

1046	Konetekniikka (K)
1051	Rakennustekniikka (R)
1052	Sähkötekniikka (S)
1055	Tuotantotalous (TU)
1137	Tietotekniikka (TI)
1144	Automaatiotekniikka (A)
1161	Materiaalitekniikka (M)
1172	Ympäristö- ja energiatekniikka (Y)
1225	Tietojohtaminen (TJ)
1231	Signaalinkäsittely ja tietoliikennetekniikka (SITI)
1260	Teknis-luonnontieteellinen (TL)
1265	Biotekniikka (B)

Tutkimuksen kohdejoukossa olevat opiskelijat jakautuivat koulutusohjelmittain kuvan 5.2 mukaisesti.

Histogrammissa on kuvattu opiskelijoiden suhteellinen osuus sekä kohdejoukkoon (118 opiskelijaa) että syksyllä 2011 kyseisessä koulutusohjelmassa aloittaneisiin opiskelijoihin (yhteensä 751 opiskelijaa). Yksikään teknis-luonnontieteellisen koulutusohjelman opiskelija ei suorittanut kurssia Insinöörimatematiikka 1u, koska kyseisillä opiskelijoilla oli pakollisena vastaava matematiikan peruskurssi Laaja matematiikka 1u. Näin ollen he eivät suorittaneet myöskään Insinöörimatematiikka 1u -kurssin tenttiä, joka toimi yhtenä muuttujana tutkmuksessa.

Kuvasta nähdään, että suhteellisesti eniten opiskelijoita on tietotekniikan koulutusohjelmasta ja pienimmät osuudet on tuotatantotalouden sekä biotekniikan koulutusohjelmista, joista kummastakin on kohdejoukossa vain yksi opiskelija. Noin 20 prosenttia automaatiotekniikan, materiaalitekniikan, tietojohtamisen sekä signaalin-

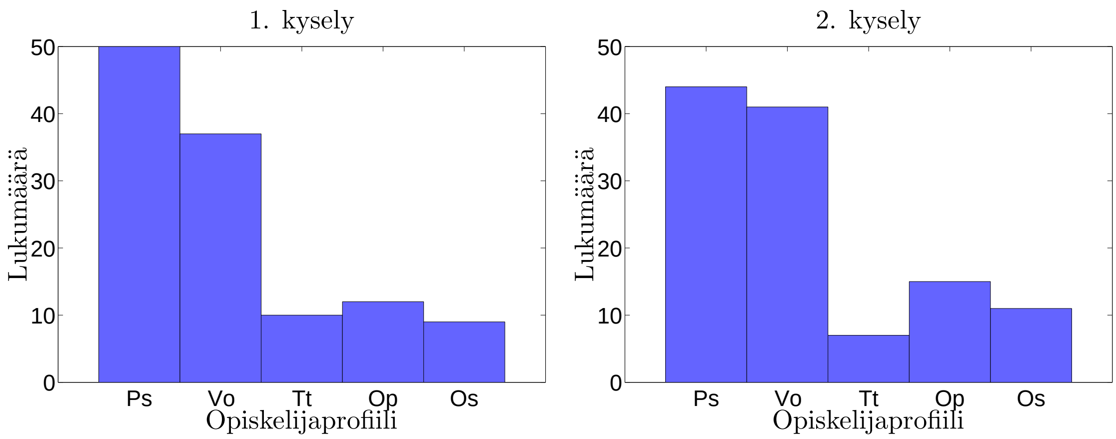


Kuva 5.2: Jumpan suorittaneiden opiskelijoiden jakautuminen koulutusohjelmittain. Vasemmanpuoleinen pylväs kuvaa suhteellista osuutta jumpan suorittaneista opiskelijoista (118) ja oikeanpuoleinen pylväs puolestaan kuvaa suhteellista osuutta koulutusohjelman koosta. Koulutusohjelmien kokona on käytetty syksyllä 2011 opintonsa aloittaneiden opiskelijoiden lukumääriä.

käsittelyn ja tietoliikennetekniikan opiskelijoista ohjattiin jumppaan ja suoritti sen hyväksytysti.

Opiskelijaprofiilit

Opiskelijalle itselleen parhaiten sopivaa opiskelijaprofiilia kuvaa kaksi luokittelumuuttujaa, sillä opiskelijaprofiilia kysyttiin sekä perustaitotestin että jälkitestin jälkeen. Erilaisia vaihtoehtoja oli viisi kappaletta ja opiskelijoille annetut vaihtoehdot sekä profiilien määrittely on esitetty aliluvussa 2.4. Kuvassa 5.3 on esitetty histogrammit sekä perustaitotestin että jälkitestin ohessa suoritetuista kyselyistä.



Kuva 5.3: Testien yhteydessä pidettyjen opiskelijaprofiilikyselyjen tulosten jakautuminen viiteen profiiliin: Pintasuuntautuneet (Ps), Vertaisoppijat (Vo), Tukea tarvitsevat (Tt), Omin päin opiskelevat (Op) sekä Osaajat (Os).

Jumpan suorittaneet opiskelijat painottuvat voimakkaasti kahteen ensimmäiseen profiiliin Pintasuuntautuneisiin mallista oppijoihin sekä Vertaisoppijoihin. Tukea tarvitsevia on suhteellisen vähän ottaen huomioon kohdejoukoon oletettua heikommat taidot matematiikassa. Toisaalta profiilikyselyn väitteiden asettelu selittää mahdollisesti edellä mainittuja asioita, sillä opiskelijoiden vastauksiin vaikuttaa mahdollisesti myös väitteiden järjestys. Tämän lisäksi yliopisto-opintonsa aloittanut tekniikan alan opiskelija harvoin tuntee tarvitsevansa tukea matematiikassa, koska vertailupohja on todennäköisesti vielä tässä vaiheessa lukio-opiskelussa. Tukea tarvitsevien määrä kuitenkin väheni matematiikkajumpan aikana, joten opiskelijat eivät heikon perustaitotestin jälkeenkään koe kuuluvansa tähän ryhmään. Taulukossa 5.4 on esitetty kaikkien perustaitotestiin osallistuneiden Insinöörimatematiikka 1u -kurssia suorittavien opiskelijoiden sekä tutkimuskohteena olevien opiskelijoiden jakaumat ensimmäisessä perustaitotestissä ja tutkimuskohteena olevien opiskelijoiden jakautuma toisessa profiilikyselyssä.

Taulukko 5.4: Opiskelijoiden jakautuminen profiilien perusteella.

	Kaikki (Prof. 1)		Jumppa (Prof. 1)		Jumppa (Prof. 2)	
Profiili	n	%	n_1	%	n_2	%
Pintasuuntautuneet	154	24,6	50	42,3	44	37,3
Vertaisoppijat	221	35,3	37	31,4	41	34,7
Tukea tarvitsevat	40	6,4	10	8,5	7	5,9
Omin päin oppijat	76	12,1	12	10,2	15	12,8
Osaajat	135	21,6	9	7,6	11	9,3
Yhteensä	626	100	118	100	118	100

Taulukosta nähdään, että pintasuuntautuneita opiskelijoita on jumpan suorittaneissa opiskelijoissa suhteellisesti selvästi enemmän kuin kaikissa perustaitotestin suorittaneissa opiskelijoissa. Toisaalta itsensä osaajaksi kokevista opiskelijoista vain pieni osa joutui matematiikkajumppaan ja läpäisi sen. Lisäksi tukea tarvitsevien osuus jumppajista on hieman koko otosta suurempi. Nämä tulokset ovat varsin ennalta odotettuja, mikä kertoo opiskelijaprofiilien toimivuudesta opiskelijoiden ryhmittelyssä.

Opiskelijaprofiilien välillä havaittiin eroaja, kun niiden toimintaa ja tuloksia vertailtiin keskenään. Taulukossa 5.5 on esitetty suhteellisten muuttujien keskiarvot opiskelijaprofiileittain perustaitotestin yhteydessä pidetyssä opiskelijaprofiilikyselyssä. Taulukosta nähdään, että opiskelijan osaamista mittaavien muuttujien keskiarvot poikkeavat eri profiilien välillä. Ensimmäisessä opiskelijaprofiilikyselyssä itsensä tukea tarvitseviksi kokeneet opiskelijat saivat jälki- ja perustaitotestissä sekä tentissä muita opiskelijaprofiileita heikompia tuloksia. Lisäksi Pintasuuntautuneet ja Vertaisoppijat saivat testeistä hieman heikompia tuloksia kuin Ominpäin opiskelevat ja Osaajat.

Taulukko 5.5: Suhteellisella asteikolla mitattavien muuttujien keskiarvot opiskelijaprofiileittain ensimmäisen profiilikyselyn perusteella.

Ensimmäinen opiskelijaprofiilikysely					
Muuttuja	Pinta-suunt.	Vertais-oppijat	Tukea tarv.	Ominp. opisk.	Osaajat
Perustaitotesti	4,44	4,81	3,40	5,17	5,33
Jälkitesti	10,92	11,38	9,60	12,75	12,22
Parannus	6,48	6,57	6,20	7,58	6,89
IMa1u	1,52	1,46	0,60	1,33	1,89
Klikkaukset	339,14	332,41	357,80	264,83	298,44
Avaukset	133,46	129,59	112,20	100,25	99,00
Yritykset	136,08	140,38	157,00	118,67	156,67
Mallivastaukset	69,60	62,43	88,60	45,92	42,78

Lokitapahtumien tuottaminen erottaa myös eri opiskelijaprofiileja toisistaan. Ensimmäisen kyselyn perusteella Tukea tarvitsevat ovat tuottaneet enemmän lokitapahtumia, eli klikkauksia, kuin muut profiilit. Vähiten klikkauksia ovat tuottaneet Ominpäin opiskelevat ja Osaajat. Ensimmäisen kyselyn perusteella Osaajat ovat pyrkineet tekemään tehtävät avaamisen jälkeen valmiiksi yrittäen ratkaista tehtävät itse ja turvautumalla harvoin mallivastauksiin. Tukea tarvitsevat ovat tuottaneet yhtä paljon ratkaisu yrityksiä kuin Osaajat, mutta tämän lisäksi he ovat availleet tehtäviä ja katsoneet malliratkaisuja useammin. Omin päin opiskelevat eroavat Osaajista pienemmässä ratkaisuyritysten määrässä.

Taulukossa 5.6 on esitetty vastaavasti suhteellisten muuttujien keskiarvot opiskelijaprofiileittain jälkitestin yhteydessä pidetyssä opiskelijaprofiilikyselyssä. Opiskelija saattoi siis vaihtaa näkemystään omasta opiskelijaprofiilistaan matematiikkajumpan aikana.

Taulukko 5.6: Suhteellisella asteikolla mitattavien muuttujien keskiarvot opiskelijaprofiileittain toisen profiilikyselyn perusteella.

Toinen opiskelijaprofiilikysely					
Muuttuja	Pinta-suunt.	Vertais-oppijat	Tukea tarv.	Ominp. opisk.	Osaajat
Perustaitotesti	4,34	4,83	3,57	5,00	5,00
Jälkitesti	10,57	11,61	11,43	12,27	11,00
Parannus	6,23	6,78	7,86	7,27	6,00
IMa1u	1,23	1,54	1,00	1,60	1,91
Klikkaukset	352,09	328,88	315,14	266,27	320,18
Avaukset	135,25	126,66	106,57	99,53	118,27
Yritykset	137,82	141,90	137,86	129,80	146,18
Mallivastaukset	79,02	60,32	70,71	36,93	55,73

Profiilien väliset erot tasoittuvat toisen kyselyn mukaisilla profiileilla. Varsinkin jälkitestin perusteella Tukea tarvitsevat eivät juurikaan eroaisi muista opiskelijaprofiileista. Insinöörimatematiikka 1u:n tenttiarvosanan suhteen ei ole tapahtunut merkittävää muutosta. Toisen profiilikyselyn perusteella ei voida tehdä samoja johtopäätöksiä profiilien opiskelutavoista kuin tehtiin ensimmäisen kyselyn yhteydessä. Omin päin opiskelevat vaikuttavat opiskelleensa vähemmin klikkauksin yrittäen ratkaista tehtävät ilman malliratkaisuja. Pintasuuntautuneet oppijat ovat tunnistaneet ensimmäistä kyselyä paremmin omat opiskelutapansa, mikä näkyy suurena mallivastausten katsomisen keskiarvona.

Opiskelijaprofiilien välillä havaittua vaihtuvuutta on kuvattu taulukossa 5.7. Rivillä on kuvattu perustaitotestin yhteydessä tehdyn ensimmäisen profiilikyselyn mukainen opiskelijaprofiili. Sarakkeet puolestaan kertovat jälkitestin yhteydessä tehdyn toisen profiilikyselyn mukaiset profiilit. Esimerkiksi ensimmäinen rivi kuvaa, mihin profiileihin ensimmäisessä profiilikyselyssä Pintasuuntautuneet mallista oppijat sijoittuivat toisessa profiilikyselyssä.

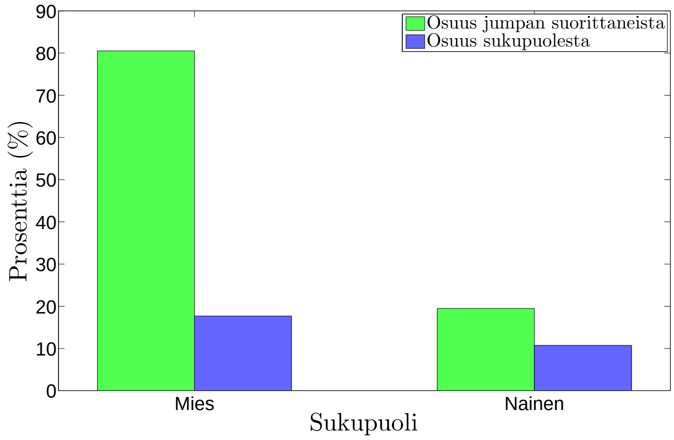
Taulukko 5.7: Opiskelijoiden profiilin vaihtuminen profiilikyselyjen välillä. Rivillä on kuvattu ensimmäisen profiilikyselyn mukainen opiskelijaprofiili ja sarakkeet kertovat toisen kyselyn mukaiset profiilit. Profiileista käytetään lyhenteitä: Pintasuuntautuneet (Ps), Vertaisoppijat (Vo), Tukea tarvitsevat (Tt), Omin päin opiskelevat (Op) sekä Osaajat (Os).

Profiili	Ps	Vo	Tt	Op	Os	Yht.
Ps	37	9	1	2	1	50
Vo	4	26	2	3	2	37
Tt	2	4	3	0	1	10
Op	1	1	1	8	1	12
Os	0	1	0	2	6	9
Yht.	44	41	7	15	11	118

Merkittävin muutos opiskelijaprofiilien välillä tapahtui Tukea tarvitsevien kohdalla. Seitsemän opiskelijaa kymmenestä koki kuuluvansa tukea tarvitsevan sijasta muihin ryhmiin. Myös yhdeksän Pintasuuntautuneista mallista oppijoista koki omaksi ryhmäkseen paremmin Vertaisoppijat. Lisääntynyt kokemus yliopisto-opiskelusta ja tietoinen omien opiskelutapojen analysointi ovat saattaneet vaikuttaa ryhmien vaihtuvuuteen. Huolimatta opiskelijaprofiilikyselyjen välisistä eroista voidaan todeta profiilikyselyn merkitys opiskelijoiden havainnoinnin kannalta. Opiskelijaprofiilit ryhmittelevät opiskelijoita melko hyvin ja esimerkiksi Tukea tarvitsevien kohdalla voitaisiin kohdistaa tarvittavia tukitoimenpiteitä, jotta pohjatiedot ja opiskelumenetelmät saataisiin yliopisto-opintojen vaatimalle tasolle.

Sukupuoli

Opiskelijoiden sukupuoli määritettiin tutkimuksessa opiskelijoiden etunimen perusteella. Tutkimuksessa haluttiin nähdä mahdolliset erot eri sukupuolten välillä, vaikka niiden tutkiminen ei ollutkaan varsinaisena tutkimuskysymyksenä. Muuttujana sukupuoli on luokittelumuuttuja, ja eri sukupuolia on aineistossa merkitty käyttämällä miehestä numeroa 1 ja naisesta numeroa 2.



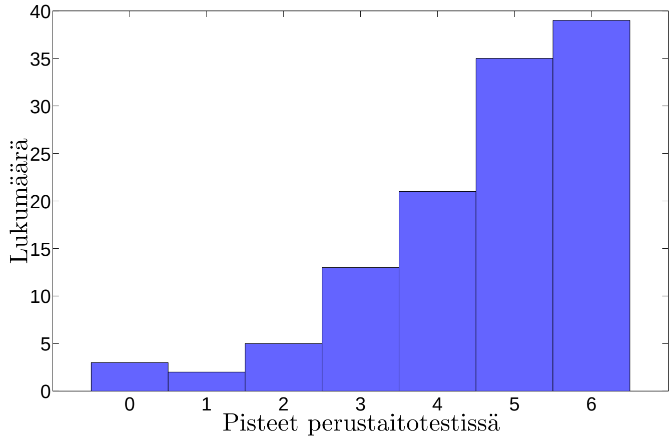
Kuva 5.4: Jumpan suorittaneiden opiskelijoiden jakautuminen sukupuolen mukaisesti. Vasemmanpuoleinen pylväs kuvaa suhteellista osuutta jumpan suorittaneista opiskelijoista (118) ja oikeainpuoleinen pylväs puolestaan kuvaa suhteellista osuutta sukupuolen koosta. Sukupuolen kokona on käytetty syksyllä 2011 opintonsa aloittaneiden opiskelijoiden tietoja.

Kuvassa 5.4 on esitetty jumpan suorittaneiden opiskelijoiden jakautuminen sukupuolen mukaisesti. Kohdejoukon opiskelijoista neljä viidesosaa oli miehiä. Kuvan avulla voidaan verrata jumpan suorittaneita opiskelijoita syksyllä 2011 aloittaneiden opiskelijoiden lukumäärään. Miehistä yli 15 prosenttia miehistä suoritti matematiikkajumpan. Naisilla vastaava prosenttiosuus oli noin 10 prosenttia.

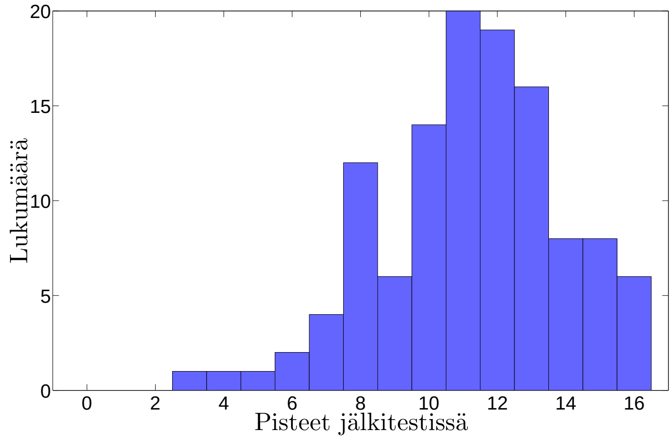
5.1.2 Perustaitotesti ja jälkitesti

Perustaitotestin ja jälkitestin tuloksia kuvaavia muuttujia kuvataan absoluuttisella asteikolla, koska selvästi jokaisella arvolla on vain yksi merkitys. Perustaitotestin tuloksissa vaihteluväli on $[0, 6]$ pistettä, koska tätä enemmän pisteitä saaneita Insinöörimatematiikan 1u:n suorittajia ei ohjattu matematiikkajumppaan.

Tutustutaan ensin muuttujien perustaitotesti, jälkitesti sekä parannus kuvaamiseen visuaalisesti. Kuvassa 5.5 on esitetty pistejakauma jumppaa edeltäneessä perustaitotestissä. Koska kyseessä on vain osa kaikista perustaitotestin suorittaneista Insinöörimatematiikka 1u:n opiskelijoista, niin histogrammi kuvaa vain vasenta reunaa suuremmasta aineistosta. Pelkän perustaitotestin perusteella ei vielä saada hyvää kuvaa opiskelijoista ja heidän ryhmittelynsä olisi vaikeaa. Toimintaa matematiikkajumpassa analysoitaessa mielenkiinto kohdistuu tämän vuoksi perustaitotestin ohella jumpan jälkeiseen jälkitestiin, joka koostui vastaavista tehtävistä kuin perustaitotestikin. Kuvassa 5.6 on esitetty jälkitestin tulokset.



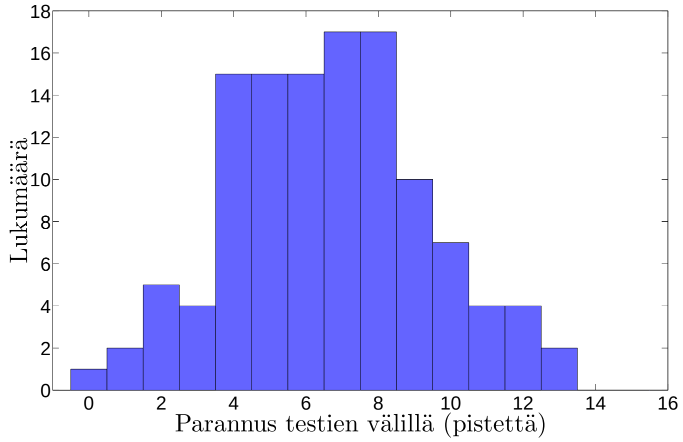
Kuva 5.5: Jumpan suorittaneiden opiskelijoiden pisteet jumppaa edeltäneessä perustaitotestissä. Maksimipistemäärä oli 16 pistettä.



Kuva 5.6: Jumpan suorittaneiden opiskelijoiden pisteet jumpan jälkeisessä perustaitotestissä (jälkitestissä). Maksimipistempäärä oli 16 pistettä.

Jälkitestissä havaitaan selkeä tulosten paraneminen. Ainoastaan viisi opiskelijaa jumpan suorittaneista ei olisi saanut perustaitotestiä hyväksytysti suoritetuksi. Matematiikkajumpasta näyttäisi olevan selvästi hyötyä matematiikan oppimisessa. Huomattava tekijä jälkitestin tuloksia analysoitaessa on kuitenkin se, että opiskelijat ovat tehneet tehtävät ilman paineita läpipääsystä sekä ilman valvontaa mahdollisten tukimateriaalien kera. Lisäksi tuloksiin vaikuttaa myös opiskelijoiden kehittynyt kyky tehdä Math-Bridge-tehtäviä. Tämä ei ole matematiikkajumpan tarkoitus, mutta perustaitotestin yksi ongelma opiskelijoilla on juuri tietokoneella tehtävien tehtävien tekeminen, johon ei ole ennen yliopisto-opiskelua totuttu.

Edellä mainitut seikat selittävät yleisesti parantuneita tuloksia testien välillä sekä jälkitestissä heikosti suoriutuneiden opiskelijoiden mahdollisia motivaatio-ongelmia suorittaa jälkitesti parhaansa mukaan. Tarkemmin testien välistä parannusta on kuvattu kuvassa 5.7, missä muuttuja parannus on erotus jälkitestin ja perustaitotestin



Kuva 5.7: Jumpan suorittaneiden opiskelijoiden parannus perustaitotestin ja jälkitestin välillä.

tuloksista. Yhdenkään opiskelijan tulokset eivät ole heikentyneet testien välillä, joka osoittaa matematiikkajumpan hyötyä sen suorittajille. Samat virhelähteet, jotka vaikuttivat jälkitestissä, vaikuttavat myös parannuksen kohdalla.

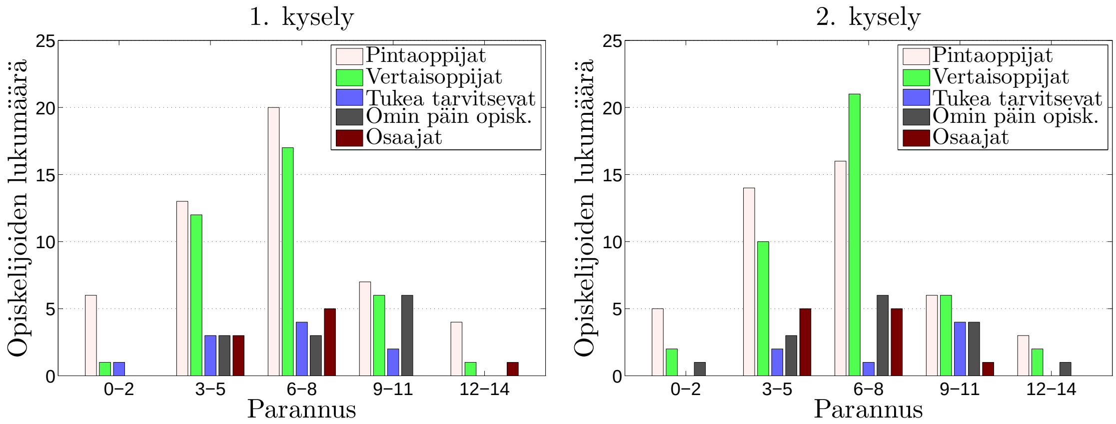
Tarkastellaan seuraavaksi muuttujien tilastollisia peruslukuja. Taulukossa 5.8 on esitetty muuttujien keskiarvo, varianssi sekä hajonta. Taulukosta nähdään, että jälkitestin ja parannuksen varianssit ovat lähellä toisiaan. Tämä selittyy perustaitotestin vähäisellä vaikutuksella parannuksen varianssiin, vaikka se vaikuttaakin keskiarvoon. Perustaitotestin vaihteluväliin nähden korkea keskiarvo ja pieni varianssi kertovat opiskelijoiden kasaantumisesta lähelle perustaitotestin läpipääsyrajaa, joka nähtiin jo kuvassa 5.5.

Taulukko 5.8: Perustaitotestin, jälkitestin sekä näiden parannuksen tilastolliset perussuureet.

Muuttuja	Keskiarvo	Varianssi	Keskihajonta
Perustaitotesti	4,61	2,10	1,45
Jälkitesti	11,24	7,14	2,67
Parannus	6,63	7,35	2,71

Seuraavaksi tarkastellaan parannusta perustaitotesteissä luokittelemalla se opiskelijaprofiilien mukaisesti. Kuvassa 5.8 on esitetty eri opiskelijaprofiilien jakautuminen parannuksen mukaisesti. Muuttujan arvoja on tiivistetty luokittelemalla ne kolmen arvon välein.

Kuvassa nähdään vasemmalla jakauma perustaitotestin yhteydessä tehdyn opiskelijaprofiilikyselyn perusteella ja oikealla jälkitestin yhteydessä tehdyn kyselyn perusteella. Vähiten parantaneista opiskelijoista suurin osa on pintasuuntautuneita tai vertaisoppijoita, kuten matematiikkajumpassa kokonaisuudessaankin. Jakaumat näillä ja tukea tarvitsevilla ovat painottuneet enemmän kuvaajan vasempaan reu-



Kuva 5.8: Opiskelijaprofiilien jakautuminen perustaitotestin ja jälkitestin parannuksen perusteella.

naan pienille parannuspisteille. Omin päin opiskelevat sekä osaajat painottuvat enemmän kuvaajan keskivaiheille tai suuremmille parannuspisteille.

Vähiten parantaneiden opiskelijoiden kohdalla profiilin muutos on tasoittanut hieman profiilien välisiä eroja. Keskivaiheille sijoittuneiden opiskelijoiden vaihtuvuus on aiheuttanut pintasuuntautuneiden ja tukea tarvitsevien oppijoiden pienenemistä ja vastaavasti vertaisoppijoiden sekä omin päin opiskelevien määrän kasvua. Suurimmilla parannusmäärillä muutokset ovat pieniä.

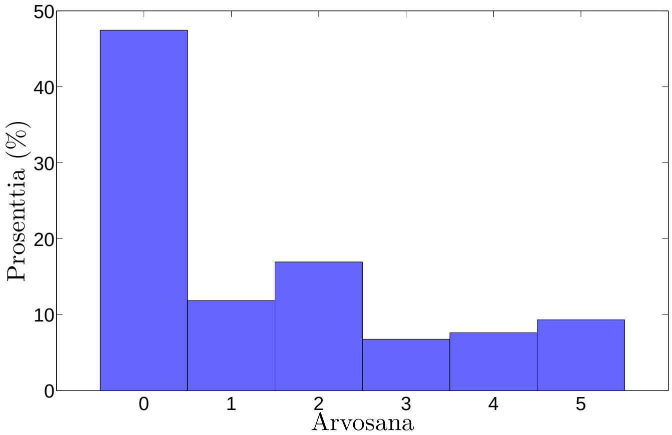
5.1.3 Insinöörimatematiikka 1u

Jumppaajien menestystä ensimmäisessä matematiikan peruskurssissa kuvataan Insinöörimatematiikka 1u -kurssin ensimmäisen tentin arvosanalla. Muuttuja on suhdeasteikollinen ja arvosanojen vaihteluväli on $[0, 5]$, jossa arvosanalla nolla tarkoitetaan hylättyä. Tällä muuttujalla voidaan tarkastella niin perustaitotestin kuin matematiikkajumpankin hyödyllisyyttä. Taulukossa 5.9 on kuvattu muuttujan tilastolliset tunnusluvut.

Taulukko 5.9: Kurssille Insinöörimatematiikka 1u osallistuneiden ja matematiikkajumpan suorittaneiden opiskelijoiden ensimmäisen tentin arvosanojen tilastolliset tunnusluvut.

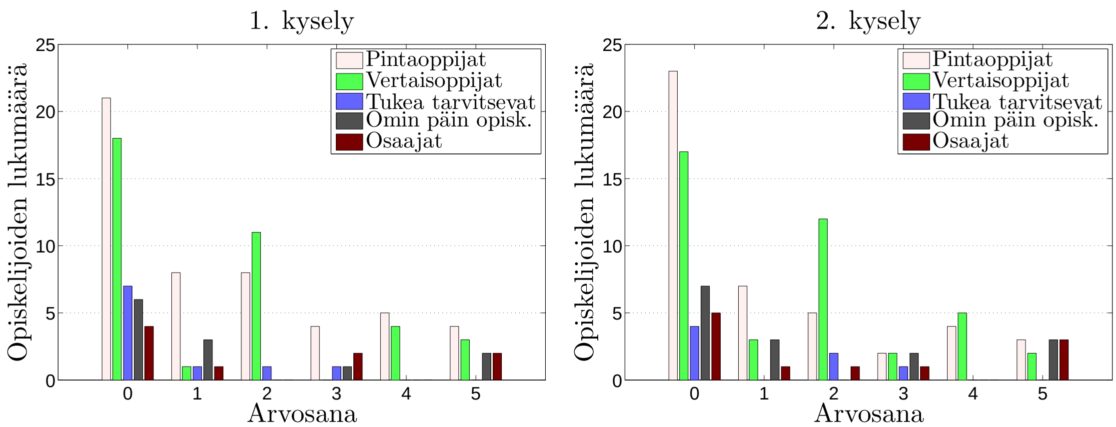
Keskiarvo	Mediaani	Varianssi	Hajonta
1,43	1	2,93	1,71

Kuvassa 5.9 on esitetty histogrammi tutkimuksen kohdejoukkoon kuuluvien 118 opiskelijan jakautumisesta heidän saamiensa arvosanojen perusteella. Kuvasta nähdään, että lähes puolet jumppaajista on saanut tentistä hylätyn arvosanan. Muut opiskelijat painottuvat myös heikoimmille arvosanoille. Kuitenkin yli 15 prosenttia opiskelijoista on saanut joko arvosanan neljä tai viisi, mikä kertoo, että matematiikkajumpassa on myös opiskelijoita, joilla ei huonosti suoritetusta perustaitotestistä



Kuva 5.9: Insinöörimatematiikan arvosanojen jakautuminen jumppaajien kesken.

huolimatta ole ongelmia matematiikan yliopisto-opintojen kanssa. Syitä huonosti menneeseen perustaitotestiin voivat näiden opiskelijoiden kohdalla olla esimerkiksi ongelmat tietokoneavusteisia tehtäviä tehdessä tai mahdolliset välivuodet, joiden aikana matematiikan perusasiat ovat unohtuneet ja jotka ovat sitten kurssin ja jumpan edetessä palautuneet mieleen. Toisaalta nähdään myös perustaitotestin hyödyllisyys kartoittaessa opiskelijoita, joilla saattaa tulla ongelmia matematiikan opinnoissa. Lisäksi havaitaan matematiikkajumpan hyödyllisyys matematiikan perusasioiden kertaamisessa, niillä opiskelijoilla, joilla taidot ovat syystä tai toisesta päässeet unohtumaan.



Kuva 5.10: Opiskelijaprofiilien jakautuminen kurssin Insinöörimatematiikka 1u tentin arvosanan perusteella.

Seuraavaksi tarkastellaan opiskelijaprofiilien jakautumista tentin arvosanojen perusteella, joka on esitetty kuvassa 5.10. Kuvassa vasemmalla on kuvaaja perustaitotestin yhteydessä ja oikealla jälkitestin yhteydessä pidetyn opiskelijaprofiilikyselyn jakautumisesta eri arvosanoille. Kuvasta nähdään, että seitsemän kymmenestä ensimmäisen kyselyn perusteella profiililtaan tukea tarvitsevista sai tentistä hylätyn arvosanan. Se on huonoin suhde kaikista profiileista, joka kertoo profiilikyselyn onnistumisesta tässä suhteessa. Hylätyn arvosanan saaneista opiskelijoista tukea tarvitsevien sekä vertaisoppijoiden määrä väheni, kun taas osaajien, omin päin opiskelijoiden sekä pintaoppijoiden määrä puolestaan kasvoi. Alimpia arvosanoja saaneet pintasuuntautuneet oppijat kokivat olevansa ennemminkin vertaisoppijoita, kun taas parhaan arvosanan saaneet opiskelijat kokivat sopivansa profiililtaan ennemmin osaajiin tai ominpäin opiskeleviin. Tentin arvosanan perusteella voidaan siis havaita jumppaajien pohtivan myös omia opiskelutapojansa.

5.1.4 Toiminta matematiikkajumpassa

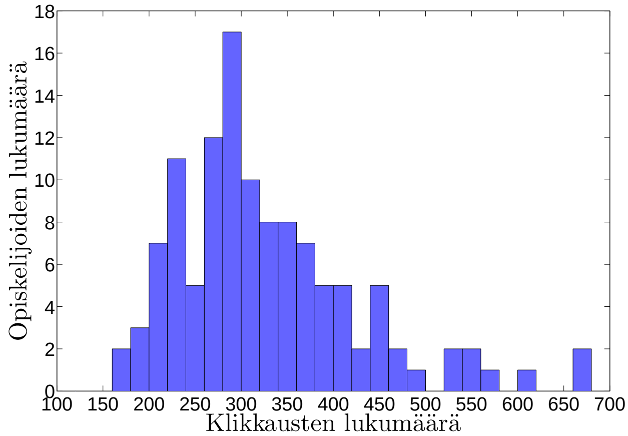
Tutkimuksen varsinaista aihetta, eli toimintaa matematiikkajumpassa lokitietojen perusteella, kuvaa erilaiset lokitietoja aiheuttaneet toiminnot, joita tässä tutkimuksessa kutsutaan klikkauksiksi. Klikkauksista saatiin tietoa opiskelijoiden tehtävien avaamisesta, ratkaisuyrityksestä sekä mallivastauksen katsomisesta. Lisäksi saatiin ratkaisuyrityksen palaute, eli oliko tehtävä mennyt oikein vai väärin. Kaikkia klikkaamiseen liittyviä muuttujia kuvataan absoluuttisella asteikolla. Taulukossa 5.10 on esitetty edellä mainittujen klikkausmuuttujien tilastollisia tunnuslukuja.

Taulukko 5.10: Klikkausten tilastollisia tunnuslukuja.

Muuttuja	Keskiarvo	Mediaani	Varianssi	Hajonta
Klikkaukset yht.	327,9	303,5	10 005,7	100,0
Avauskerrat	124,4	109,0	2862,7	53,5
Vastausyritykset	139,0	130,0	1534,9	39,2
Mallivastaukset	64,5	56,0	2311,5	48,1
Oikeat vastaukset	75,2	73,0	74,7	8,6
Väärät vastaukset	63,8	54,5	1456,5	38,2

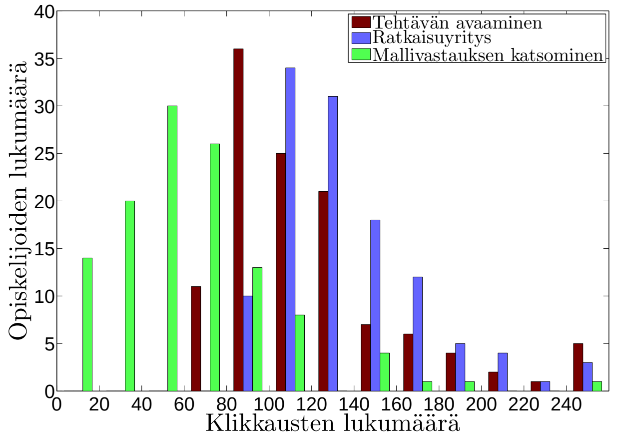
Taulukosta nähdään, että klikkausten määrä eri opiskelijoilla on hajonnut melko voimakkaasti. Toisaalta opiskelijat eivät ole ratkaisseet samaa tehtävää useampaa kertaan, sillä oikeiden vastausten keskiarvo ja mediaani ovat lähellä tehtävien lukumäärää 71 ja hajonta on pieni. Tästä johtuen vastausyritysten ja väärien vastausten hajonta ovat lähellä toisiaan.

Tarkastellaan seuraavaksi tarkemmin taulukon muuttujia. Kuvassa 5.11 on kuvattu opiskelijoiden jakautuminen kaikkien klikkausten perusteella. Aineistoa on tiivistetty luokittelemalla klikkausten määrä 20 välein. Väli on alarajalla suljettu ja ylärajalla avoin. Pienin mahdollinen klikkausten määrä on 71 tehtävän avaamista ja 71 oikein mennyttä ratkaisuyritystä eli yhteensä 142 klikkausta. Kuvasta nähdään, että yksikään opiskelija ei ole suorittanut jumppaa alle 160 klikkauksella. Eniten opiskelijoita on luokassa 280–299, ja keskimäärin opiskelijat käyttivät noin 4,6 klikkausta jokaista tehtävää kohti.



Kuva 5.11: Opiskelijoiden jakautuminen klikkausten kokonaismäärän perusteella.

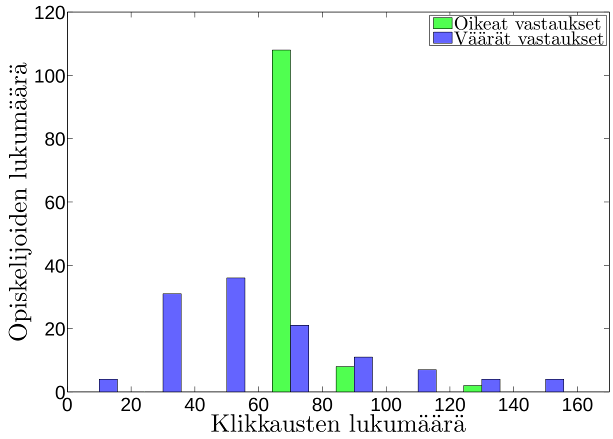
Klikkaukset voidaan jakaa tarkemmin kolmeen osaan. Tehtävien avauskertoihin, ratkaisuyrityksiin sekä mallivastausten katsomiseen. Kuvassa 5.12 on esitetty näiden muutujien jakautuminen. Aineistoa on tiivistetty luokittelemalla klikkausten määrä 20 välein. Väli on alarajalla suljettu ja ylärajalla avoin. Lisäksi viimeisessä luokassa ovat kaikki opiskelijat, joilla oli 240 klikkausta tai enemmän.



Kuva 5.12: Opiskelijoiden jakautuminen tehtävien avaamisen, ratkaisuyritysten sekä mallivastausten katsomisen perusteella.

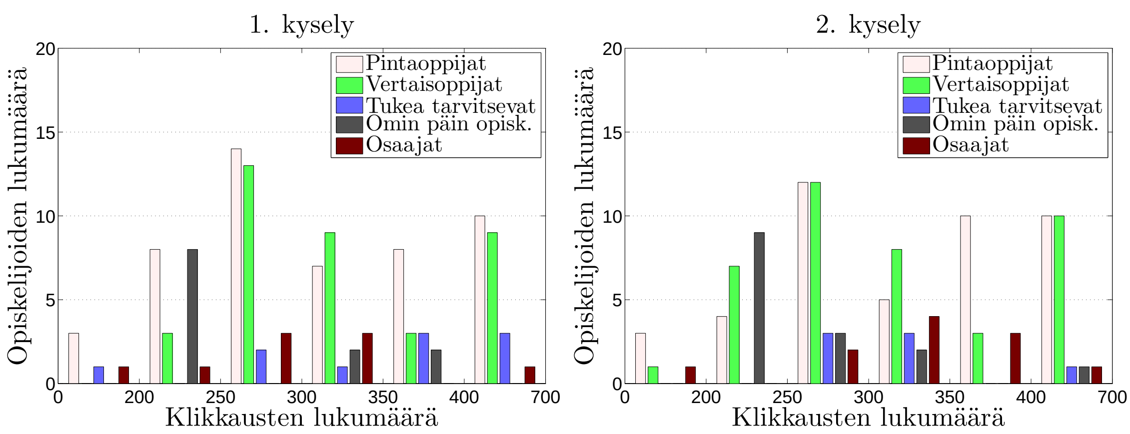
Kuvasta nähdään, että keskimäärin tehtäviä on yritetty ratkaista useammin kuin kerran sen avaamisen jälkeen, mikä oli odotettavissa. Osa opiskelijoista sai tehtävät suoritettua vain vähäisellä turvautumisella malliratkaisuihin. Tehtävien avauskertojen lukumäärän yli kolminkertainen määrä tehtävien määrään nähden saattaa kertoa opiskelijan kärsivällisyydestä tehdä avattu tehtävä valmiiksi epäonnistuneista ratkaisuyrityksistä huolimatta. Lisäksi toistuva mallivastausten katsominen ja saman mallivastauksen katsominen useampaan kertaan saattaa kertoa ongelmista keskittyä tehtävien tekemiseen.

Tehtävien ratkaisuyritykset voidaan jakaa oikeisiin ja vääriin ratkaisuihin. Kuten taulukosta 5.10 nähtiin, oikeiden vastausten määrä ei pääsääntöisesti poikkea tehtävien määrästä. Kuvassa 5.13 on kuvattu opiskelijoiden jakautuminen oikeiden ja väärin vastausten perusteella. Aineistoa on tiivistetty luokittelemalla klikkausten määrä 20 välein. Väli on alarajalla suljettu ja ylärajalla avoin.



Kuva 5.13: Oikeat ja väärät vastaukset.

Tutkimusjoukon 118:sta opiskelijasta 36 suoritti tehtävät vain 71 oikealla vastauksella tehtävää kohden. Kuvasta nähdään, että lähes kaikki opiskelijat suorittivat jumpan alle 80 oikealla ratkaisulla. Tämä ei kerro juurikaan opiskelijoiden opiskelutavoista, vaan osoittaa ettei matematiikkajumpaa käytetty yleisesti matematiikan opiskeluun vaan ohjattuun perusasioiden kertaamiseen. Väärien vastausten lukumäärä on jakautunut oikeiden vastausten hajonnasta johtuen lähes samalla tavalla kuin ratkaisuyritystenkin lukumäärä.



Kuva 5.14: Klikkausten jakautuminen profiileittain.

Tarkastellaan lopuksi klikkausten jakautumista eri opiskelijaprofiilien suhteen. Kuvassa 5.14 on esitetty histogrammi, jossa klikkausten kokonaismäärät on esitetty vasemmalla perustaitotestin yhteydessä pidetyn opiskelijaprofiilikyselyn perusteella ja oikealla jälkitestin yhteydessä pidetyn kyselyn perusteella. Aineistoa on tiivistetty luokittelemalla klikkausten määrä 50 välein. Väli on alarajalla suljettu ja ylärajalla avoin. Erityisesti jaon alku on välillä $[0, 200)$ ja loppu välillä $[400, 700]$.

Kuvasta nähdään, että suurin osa omin päin opiskelevista on klikannut alle 250 kertaa. Pintasuuntautuneista mallista oppijoista sekä vertaisoppijoista huomattavan suuri osa on puolestaan keskittynyt suuremmille klikkausmäärille. Paljon klikanneista opiskelijoista, jotka kokivat ensimmäisessä kyselyssä olevansa tukea tarvitsevia, lähes kaikki vaihtoivat profiilia toisessa kyselyssä. Osaajat ovat painottuneet hieman suuremmille klikkausmäärille.

5.1.5 Korrelaatio

Tutkimuksen alkuvaiheessa muuttujien välisiä suhteita oli helppoa lähteä tutkimaan laskemalla muuttujien välisiä korrelaatioita. Korrelaation avulla voidaan nähdä lineaarista riippuvuutta muuttujien välillä ja näin ollen yrittää selittää opiskelijoiden käyttäytymistä matematiikkajumpassa. Taulukossa 5.11 on kuvattu muuttujien välisiä korrelaatioita matriisimuodossa. Korrelaatio vapausasteilla $118 - 2 = 116$ on tilastollisesti “melkein merkitsevä” (riskitaso 5 %), jos $|r| \geq 0,181$; merkitsevä (riskitaso 1 %), jos $|r| \geq 0,236$ ja erittäin merkitsevä (riskitaso 0,1 %), jos $|r| \geq 0,299$. Lisäksi on hyvä huomioida lähtökohtaiset riippuvuudet muuttujien välillä. Parannus on jälki- ja perustaitotestin erotus, klikkaukset on summa avaus-, yritys- ja mallivastauskerroista ja yrityskerrat on summa oikeista ja vääristä vastauksista. Taulukkon korrelaatiot ovat Pearsonin tulomomenttikorrelaatiokertoimia. Muuttujille laskettiin myös Spearmanin järjestyskorrelaatiokertoimet, jotka erosivat taulukoiduista tietyiltä osin. Näistä eroavaisuuksista on mainittu muuttujakohtaisesti tarpeen mukaan.

Taulukon toisistaan ennalta riippumattomien muuttujien korrelaatiot ovat melko pieniä, mutta osa niistä on tilastollisesti merkittäviä. Mielenkiintoisia korrelaatioita ovat muun muassa mallivastausten katsomisen negatiivinen korrelaatio jälki- ja perustaitotestin kanssa. Tämä tarkoittaisi, että opiskelijat, jotka menestyvät testeissä huonosti, joutuvat jumppaa tehdessään turvautumaan malliratkaisuihin muita opiskelijoita useammin. Spearmanin vastaavat korrelaatiot mallivastausten katsomisen kanssa ovat $\rho = -0,20$ (perustaitotesti) ja $\rho = -0,25$ (jälkitesti). Nämä ovat selvästi Pearsonin kertoimia pienemmät, mutta kuitenkin merkitseviä. Kolmas opiskelijan taitotasoa mittaava testi on kurssin Insinöörimatematiikka 1u tentin tulos, joka korreloi myös negatiivisesti mallivastausten kanssa. Tämä korrelaatio korostuu käytettäessä Spearmanin korrelaatiota, jolloin $\rho = -0,28$. Malliratkaisujen katsomisen kanssa korreloi lisäksi oikeat ja väärät vastaukset sekä ratkaisuyritykset, joka on oikeiden ja väärien vastausten summamuuttuja. Näiden korrelaatioiden suuruus korostuu huomattavasti käytettäessä Spearmanin korrelaatiokerrointa. Tut-

Taulukko 5.11: Korrelaatiomatriisi muuttujille perustaitotesti (PTT), jälkitesti (JäT), Insinöörimatematiikka 1u:n tentti (IMA), klikkaukset yhteensä (Kli), avauskerrat (Ava), ratkaisuyritykset (Yri), mallivastaukset (Mal), oikeat (Oik) ja väärät vastaukset (Vää). Tähdet luvun perässä merkitsevät merkitsevyystasoa: * tarkoittaa melkein merkitsevää, ** merkitsevää ja *** erittäin merkitsevää.

	PTT	JäT	Par	IMa	Kli	Ava	Yri	Mv
PTT	1,00							
JäT	0,24**	1,00						
Par	−0,29**	0,85***	1,00					
IMa	0,10	0,30**	0,24**	1,00				
Kli	−0,22*	−0,31***	−0,19*	−0,24**	1,00			
Ava	−0,03	−0,06	−0,04	−0,06	0,72***	1,00		
Yri	−0,08	−0,26**	−0,21**	−0,24**	0,67***	0,20*	1,00	
Mav	−0,36***	−0,37***	−0,18	−0,24*	0,74***	0,22*	0,36***	1,00
Oik	−0,16	−0,10	−0,01	−0,10	0,31***	0,17	0,23**	0,27**
Vää	−0,05	−0,25**	−0,22**	−0,23*	0,62***	0,16	0,98***	0,31***

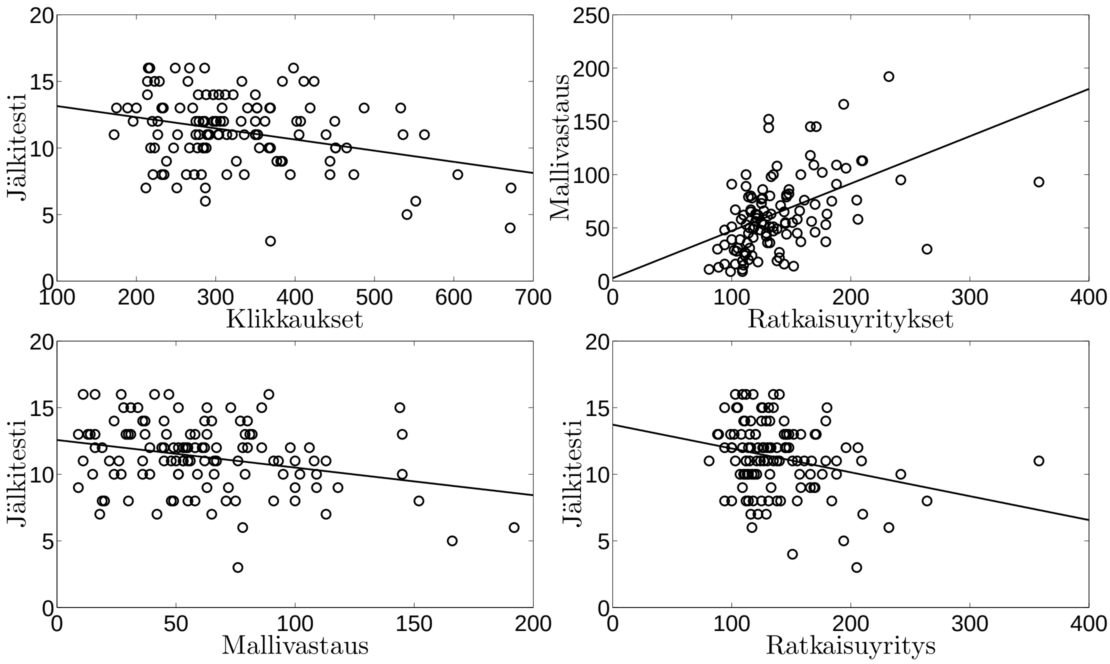
kimuksen kannalta huomioitavaa ovat suuret korrelaatiot erilaisten lokitietoja tuottaneiden muuttujien välillä, jotka korostuvat entisestään Speramanin menetelmää käytettäessä.

Matematiikkajumpan hyödyllisyydestä opiskelijoille kertoo jälkitestin positiivinen korrelaatio Insinöörimatematiikka 1u:n tentin arvosanaan. Toisaalta sekä jälki- että perustaitotesti korreloivat negatiivisesti klikkausten kokonaismäärän kanssa. Vaikka korrelaatiot ovatkin pienempiä käytettäessä Spearmanin menetelmää, niin ne kertovat kuitenkin matematiikkajumpan lokitietojen paljastavan epävarman opiskelijan, joka harhailee tehtävästä toiseen, yrittää ratkaista tehtäviä, katsoo paljon mallivastauksia ajattelematta niitä kunnolla ja lopulta luovuttaa helposti siirtyen uuteen tehtävään syventymättä mihinkään yksittäiseen tehtävään päämäärätietoisesti. Tämä näkyy myös peruskurssin tentin negatiivisena korrelaationa kaikkien klikkaamiseen liittyvien muuttujien kanssa.

Korrelaatiolla saadaan kuitenkin esiin vain suoraviivainen muuttujien riippuvuus, joka ei välttämättä näytä muuttujien välisiä yhteyksiä oikein. Tästä johtuen seuraavaksi tarkastellaan neljää muuttujaparia niiden sirontakuvioiden sekä regressiosuorien avulla.

Kuvassa 5.15 on esitetty neljä sirontakuviota ja regressiosuoraa. Vasemmalta ylhäältä alkaen, ensimmäisessä kuvaajassa nähdään jälkitestin arvosanojen jakautuminen klikkausten perusteella. Alle 500 klikkauksen määrällä ei ole merkittävää suhdetta jälkitesetin tulokseen. Vasta suuremmilla klikkausmäärillä alkavat paljastua opiskelijat, joiden jälkitestin tulokset eivät parane huolimatta lisääntyneestä aktiivisuudesta matematiikkajumpassa.

Toisessa kuvaajassa nähdään ratkaisuyritysten ja mallivastauksen katsomisen si-



Kuva 5.15: Sirontakuviot ja regressiosuorat muuttujapareille klikkaukset – jälkitesti, ratkaisuyritykset – mallivastaukset, mallivastaukset – jälkitesti sekä ratkaisuyritykset – jälkitesti.

rontakuvio. Tässä nähdään jonkinlainen yhteys muuttujien välillä, sillä mallivastausten katsominen lisääntyy ratkaisuyritystenkin lisääntyessä. Tämä on toisaalta ymmärrettävää, koska vaikeudet jumpan tehtävissä aiheuttavat molempien muuttujien kasvamista. Toisaalta tämä paljastaa ne opiskelijat, jotka ennemmin kokeilevat tehtävien tekemistä malliratkaisujen avulla kuin opiskelisivat asian kerralla riittävän hyvin. Toinen syy tällaiseen käyttäytymiseen saattaa olla opiskelijalle entuudestaan tuntematon syntaksi, jolla tehtävien ratkaisut Math-Bridgeen syötetään.

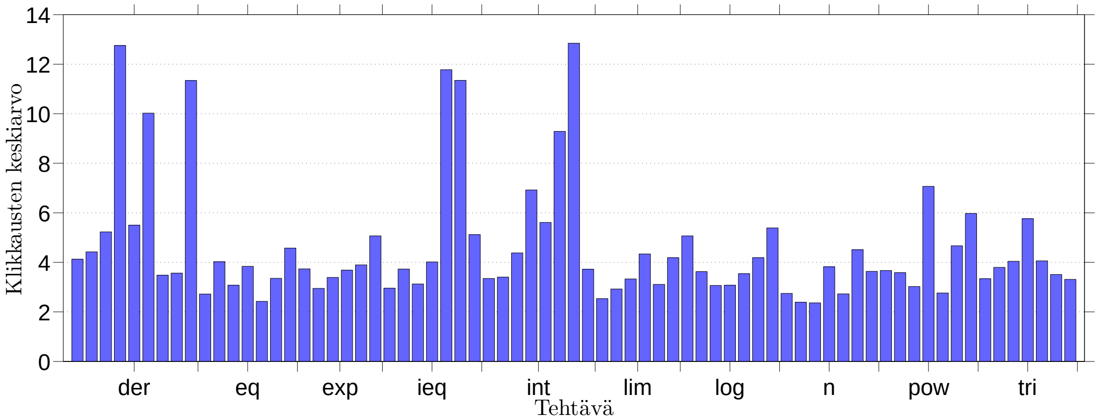
Kolmannessa kuvaajassa on esitetty jälkitestin tulosten jakautuminen mallivastausten katsomisen perusteella. Kuvaaja näyttää suuresta korrelaatiosta huolimatta erittäin hajonneelta, eikä muuttujien välillä näyttäisi olevan riippuvuutta. Kuvasta voidaan nähdä päämäärättömästi opiskelevat opiskelijat, kuten nähtiin ensimmäisestäkin kuvasta.

Neljännessä kuvaajassa on esitetty jälkitestin tulosten jakautuminen ratkaisuyritysten perusteella. Vaikka pistejoukko on hajonnut melko paljon, niin siinä on kuitenkin havaittavissa jälleen ne opiskelijat, joilla jälkitesti ei ole onnistunut monista ratkaisuyrityksistä huolimatta.

5.2 Tehtävät

Seuraavaksi tarkastellaan matematiikkajumpan tehtäviä, joiden analysoinnissa voidaan myös hyödyntää lokitietoja. Lokitietojen perusteella saadaan tehtäväkohtaiset avauskerrat, ratkaisuyritykset sekä mallivastausten katsomiset kuten opiskelijoidenkin kohdalla tehtiin. Tehtävien analysointi antaa tietoa mahdollisista vaikeista tehtävätyypeistä, joiden parissa opiskelijoilla on ollut ongelmia aiheen vaikeuden vuoksi tai muista, kuten Math-Bridgen syntaksin aiheuttamien ongelmien vuoksi.

Verrataan opiskelijoiden keskimääräistä klikkausmäärää tehtäväkohtaisesti, joka on esitetty kuvassa 5.16. Tehtävät on jaoteltu tehtäväkoodin mukaisesti aakkosjärjestyksessä. Kaikkien tehtävien tyypit on kuvailtu luvussa 3 sivulla 14 ja kaikkien tehtävien tarkemmat muodot on nähtävissä liitteessä C.



Kuva 5.16: Tehtäväkohtainen klikkausten keskiarvo matematiikkajumpassa.

Kuvaajaa tarkasteltaessa havaitaan joidenkin tehtävien aiheuttaneen selvästi enemmän klikkauksia kuin muut. Pakollisia klikkauksia tehtävää kohden on kaksi: tehtävän avaaminen ja tehtävän ratkaiseminen onnistuneesti. Tarkastellaan seuraavaksi tarkemmin seitsemää tehtävää, joita opiskelijat klikkasivat keskimäärin yli kahdeksan kertaa matematiikkajumpan aikana. Tehtävät ovat vain esimerkkejä Math-Bridgen tehtävistä, koska tehtävää avatessa tehtävän kokonaislukuvakiot generoituvat satunnaisesti.

Järjestyksesään neljäs derivointitehtävä aiheutti keskimäärin yli kaksitoista klikkausta opiskelijoilta matematiikkajumpan aikana. Tehtävän koodi on *der06*. Tehtävänä on laskea seuraavaa muotoa olevan trigonometrisen funktion derivaatta:

$$f(x) = 6\cos(9x) + 5\cos^9(6x) + 7\sin^3(5x) + 5\sin(5x).$$

Vaikka tehtävä onkin vaatimuksiltaan soveltuva lukiomatematiikan kertaustehtäväksi, vaatien useamman kuin yhden derivointimenetelmän osaamista, niin mekaaninen ratkaisu on työläs ja virheiden tekemisen määrä tehtävän syntaksia kirjoittaessa on todennäköistä. Tehtävän ratkaisu on

$$f'(x) = -54\sin(9x) - 270\cos^8(6x)\sin(6x) + 105\cos(5x)\sin^2(5x) + 25\cos(5x),$$

jolloin opiskelijan syöttämä syntaksi kahdelle riville jaettuna on

```
-54*sin(9*x)-270*cos(6*x)^8*sin(6*x)
+105*cos(5*x)*sin(5*x)^2+25*cos(5*x).
```

Syntaksin tarkastamisen mahdollisuus auttaa tällaisessa tilanteessa, mutta vastaus on silti pitkä. Tehtävä voitaisiin mahdollisuuksien mukaan jakaa useampaan osaan, jolloin tehtävien teon mielekkyys kasvaisi ilman luopumista opetuksellisista tavoitteista.

Myös kuudes derivointitehtävä, koodiltaan *der08*, sai keskimäärin noin kymmenen klikkausta jumpan suorittaneilta opiskelijoilta. Tehtävänä on derivoida logaritmifunktio, joka on muotoa:

$$f(x) = \ln\left(\frac{x-8}{x^2+6}\right).$$

Tämänkin tehtävän ratkaisemiseksi on yhdistettävä useampia kuin yhtä derivointisääntöä, mikä osaltaan vaikeuttaa tehtävää. Toisaalta se on kuitenkin tehtävänä hyvin kattava luonnollisen logaritmin, osamäärän sekä sisäfunktion derivaatan osaamisen harjoittelussa. Ratkaisun syntaksista tulee supistamisesta riippuen vaikeahko, mikä saattaa kasvattaa virheiden esiintymistä.

Derivoimistehtävistä viimeinen, *der11*, sai opiskelijoilta keskimäärin yli 11 klikkausta. Tehtävässä on yhdistetty eksponenttifunktion sekä logaritmifunktion derivointia ja tehtävän funktio on muotoa:

$$f(x) = e^{-\frac{1}{4}x}\big(\ln(9x)\big)^{\frac{1}{3}}.$$

Tehtävässä joudutaan käyttämään vähemmän tuttuja derivoimissääntöjä, joita kuitenkin aikaisemmissa tehtävissä harjoiteltiin. Tässä tehtävässä oli ongelmia myös derivoitavan funktion piirtymisessä selaimen ikkunassa, koska joissain tapauksissa eksponentissa olevat murtoluvut vaativat selaimen zoomaustyökalun käyttöä nimittäjän arvon varmistamiseksi. Tämä ei sinällään lisää virheitä ja sen myötä klikkauksia, mutta aiheuttaa tekijän kannalta ylimääräistä turhautumista. Ratkaisun syntaksin pituus aiheuttaa tässäkin tilanteessa ylimääräisiä virheitä, sillä vastaus on muotoa

$$f'(x) = -\frac{1}{4}e^{-\frac{1}{4}x}\cdot\ln(9x)^{\frac{1}{3}} + e^{-\frac{1}{4}x}\cdot\frac{1}{3x\big(\ln(9x)\big)^{\frac{2}{3}}}.$$

Derivointitehtävät vaikuttivat siis ongelmallisilta ainakin vaikeampien derivoimissääntöjen kohdalla ja näitä sääntöjä yhdisteltäessä. Näiden lisäksi kuvasta 5.16 nähdään, että viides epäyhtälötehtävä koodilla *ieq6*, sai keskimäärin lähes kaksitoista klikkausta. Tehtävässä piti ratkaista rationaaliepäyhtälö, joka on muotoa

$$\frac{7x}{5} < \frac{1-5x}{x-7},$$

ja vastauksessa piti antaa arvot kolmelle parametrille a, b ja c, jotka saadaan epäyhtälöistä $x < a$ ja $b < x < c$. Klikkaukset selittyvät tehtävän pituudella, vaikeudella ja virhealttiilla rationaalilausekkeilla, minkä vuoksi myös syntaksista tulee hankalahko. Tehtävässä opiskelija joutuu käyttämään merkkikaaviota ja suorittamaan muutenkin paljon päässälaskusta poikkeavaa laskentaa.

Myös kuudes epäyhtälötehtävä tuotti keskimäärin yli 11 klikkausta jumpan suorittaneilta opiskelijoilta. Tehtävän koodi on *ieq7*, ja siinä tulee ratkaista itseisarvoepäyhtälö, johon on yhdistetty edellisen tehtävän tapaan murtolauseke. Tehtävä oli muotoa

$$\left|\frac{3x+1}{7x-1} - \frac{3}{7}\right| < 21.$$

Tehtävä on ratkaisumalliltaan samanlainen kuin edellinen epäyhtälötehtävä, mutta nyt itseisarvojen vuoksi jouduttin ratkaisemaan kaksi epäyhtälöä, mikä tuo tehtävään lisää haastavuutta ja aiheuttaa virheitä. Tehtävän *ieq6* ratkaistuaan opiskelija osaa ratkaista tehtävän *ieq7*, joka on saattanut johtaa siihen, ettei klikkausten määrä ole noussut suuremmaksi pidemmästä ratkaisusta huolimatta.

Loput vaikeaksi osoittautuneet tehtävät olivat integrointitehtäviä. Kuudetta integrointitehtävää, *int6*, klikattiin keskimäärin yli yhdeksän kertaa. Tehtävässä piti laskea määrätty integraali polynomifunktiosta, eli esimerkiksi

$$\int_{-5}^{2} 2x^4 - 2x^2 - 3 \ dx.$$

Tämän tehtävän ei pitäisi olla millään tavalla vaikea. Polynomifunktion integroiminen on helpoin mahdollinen integrointitehtävä, eikä määrätyn integraalin laskeminen vaikeuta tehtävää merkittävästi. Lokitietojen perusteella (liite C) tehtävä ei ole aiheuttanut mallivastauksen katsomisia keskimääräistä enempää vaan lähinnä uusien lukuarvojen generoimista ja tehtävän ratkaisuyrityksiä. Aikaisemmissa integrointitehtävissä on ollut polynomifunktion integraalifunktion määrittämistä, joten todennäköisin syy on jonkinasteiset huolimattomuus- ja etumerkkivirheet integrointirajoja sijoittaessa ja vastausta laskiessa.

Seuraava integroimistehtävä, *int7*, tuotti keskimäärin noin kolmetoista klikkausta, joka oli yksi suurimmista keskiarvoista. Tehtävässä piti laskea itseisarvofunktion määrätty integraali, eli lauseke muotoa

$$\int_{-3}^{2} -3\left|x^2+x\right| dx.$$

Tehtävän alussa pitää tarkastella itseisarvomerkkien sisällä olevan sisäfunktion kulkua ja määrittää integraali uudelleen jakamalla integraali pienempiin osiin. Tässä tehtävässä ei syntaksista tule millään tavalla ongelmallista, vaan ongelmat ovat opiskelijoiden ratkaisutaidoissa. Opiskelijat katsoivat keskimäärin yli neljä kertaa mallivastausta (liite C), mikä antaa kuitenkin kattavan kuvan tehtävän vaatimasta ratkaisumallista. Tämä kertoo opiskelijoiden keskittymiskyvyn puutteesta ja rajaa melko tarkasti puutteellisia osa-alueita, joiden harjoitteluun tulee panostaa jatkossa.

Näiden tehtävien ja kuvan 5.16 perusteella nähdään suurimmat puutteet ja vaikeimmat osa-alueet opiskelijoiden matematiikan osaamisessa. Karkeasti katsottuna suurimmat ongelmat ovat derivoinnissa ja integroinnissa sekä epäyhtälöiden ja trigonometrian alueella. Tämän lisäksi lokitiedoista saadaan tietoa matematiikkajumpan tehtävien sekä Math-Bridge-ohjelmiston kehittämiseen. Matematiikkajumpan osalta tehtävien mielekkyyttä voidaan kehittää, mutta kattavaa asioiden kertaamista ja uudelleen opettelua varten tärkeämpää on saada opiskelijoiden opiskelutavat johdonmukaisemmaksi, jotta tehtävät ratkaistaan ajatuksella ja tehtävän vaatima teoria opetellaan huolellisesti ennen tehtävän ratkaisemista. Tämä vähentäisi klikkausten määrää, kun oman onnensa koettelu vaihtuisi määrätietoiseksi tehtävän ratkaisemiseksi.

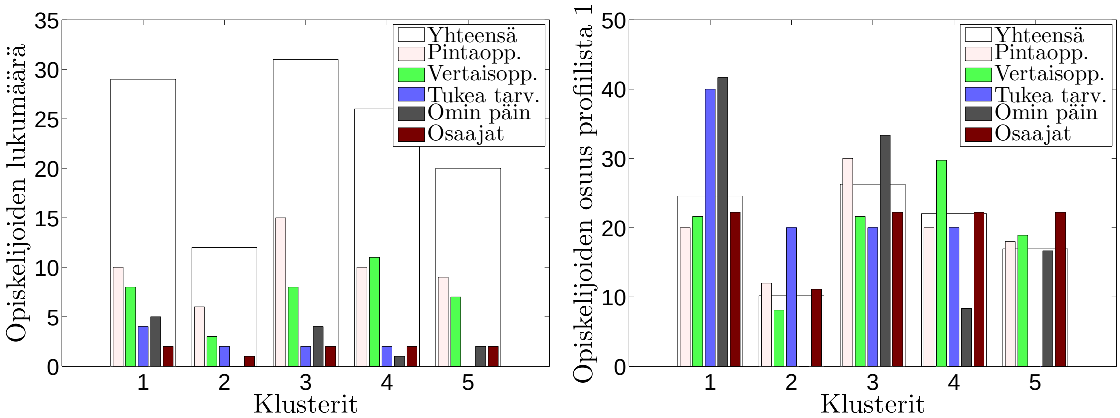
5.3 Klusterointi

Klusterointi suoritettiin aineistolle käyttämällä kolmea muuttujaa: parannus perustaitotesteissä (“Parannus”), Insinöörimatematiikka 1u:n arvosana (“IMa1u”) sekä klikkausten kokonaismäärä (“Klikkaukset”). Muuttujien valinnassa käytettiin sivulla 56 olevan taulukon 5.11 korrelaatioita. Niistä havaittiin, että kaikki erilaiset lokitapahtumat korreloivat voimakkaasti kokonaisklikkausten määrän kanssa, joten se on sopiva muuttuja kuvaamaan toimintaa matematiikkajumpassa.

Tämän lisäksi haluttiin saada toisistaan riippumattomat oppimistuloksia kuvaavat muuttujat. “Parannus” kuvaa hyvin tietokoneella tehtävien Math-Bridge-tehtävien suorittamista ja opiskelijan kehittymistä niiden pohjalta. Insinöörimatematiikka 1u:n arvosana puolestaan kuvaa opiskelijan suoriutumista matematiikkajumpan ulkopuolisesta testistä, johon matematiikkajumpalla pyritään myös antamaan valmiuksia.

Klustereiden määräksi valikoitui viisi, koska vähemmällä määrällä klustereita ei saatu riittävästi eroteltua opiskelijoita. Toisaalta suurempi määrä klustereita olisi aiheuttanut liikaa samankaltaisia ryhmiä, jolloin niiden analysointi, ominaispiirteiden havaitseminen sekä nimeäminen ei olisi ollut mielekästä.

Klusterointiin käytettiin K-means klusterointia. Aineistoa klusteroitaessa hierarkisella klusteroinnilla saatiin aina hyvin erikokoisia klustereita ja usein vain yhden opiskelijan muodostamia klustereita. Tämä kertoo joidenkin opiskelijoiden selvästä eroavaisuudesta muihin opiskelijoiden nähden, kun käytetään kyseisiä muuttujia. K-means klusterointi antoi kuitenkin järkevämmät ja tulkittavammat klusterit, kun yhden opiskelijan vaikutus klusterikeskukseen ei ollut liian suuri. Klusterointi suoritettiin Matlab-ohjelmiston `kmeans`-funktion avulla ja algoritmi suoritettiin 20 kertaa eri alkuarvoilla, jolloin saatiin varmuudella paras mahdollinen klusterointi.



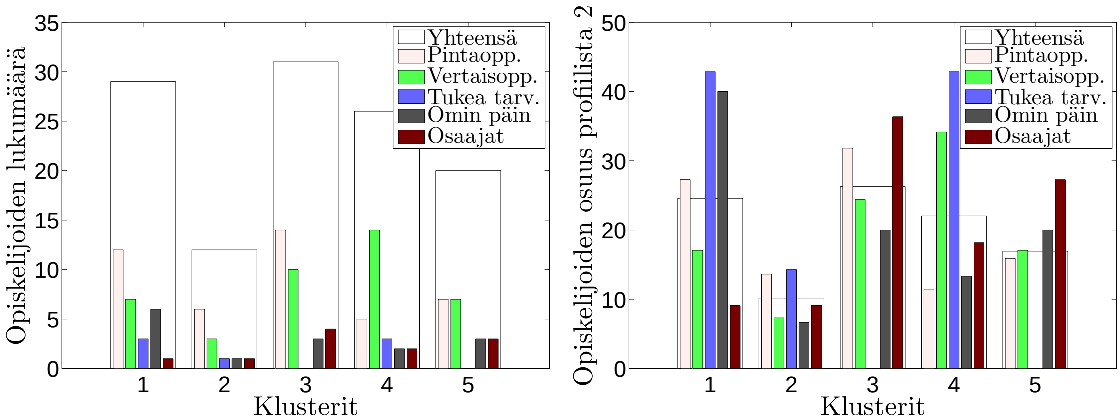
Kuva 5.17: Ensimmäisen profiilikyselyn mukaisten opiskelijaprofiiliryhmien jakautuminen klustereihin. Vasemmalla pylväät kuvaavat lukumäärää ja oikealla prosenttiosuutta opiskelijaprofiilin koosta.

Verrataan klustereiden tuloksia ensin opiskelijaprofiileihin. Kuvassa 5.17 on esitetty klusterien jakautuminen ensimmäisen profiilikyselyn perusteella. Vasemmanpuoleisessa kuvassa nähdään valkoisella palkilla kuvattuna opiskelijoiden kokonaisfrekvenssit klustereissa, jotka on numeroitu yhdestä viiteen. Kapeammat palkit kuvaavat eri opiskelijaprofiileita. Jokainen kapea palkki on niiden opiskelijoiden frekvenssi, jotka kuuluvat kyseiseen profiiliin ensimmäisen kyselyn perusteella.

Oikeanpuoleisessa kuvassa on kuvattu vastaava frekvenssi suhteellisena prosenttiosuutena taustalla olevan valkoisen palkin tapauksessa kaikista jumppaajista ja kapeampien palkkien kohdalla on laskettu suhde kyseisen profiilin kokoon. Toisin sanoen kuvaajasta nähdään, kuinka monta prosenttia kustakin profiilista kuuluu tiettyyn klusteriin. Tällaisen kuvaajan avulla on helpompi nähdä erikokoisten profiilien kohdalla suhteelliset jakaumat eri klustereiden välillä. Kuvassa 5.18 on vastaavalla tavalla esitetty histogrammit jälkitestin yhteydessä pidetyn opiskelijaprofiilikyselyn perusteella.

Kuvista nähdään, että opiskelijoiden jakautuminen klustereihin on melko satunnaista. Toinen klusteri on pienin, alle 15 opiskelijan, klusteri, kun muissa klustereissa on 20 opiskelijaa tai enemmän.

Hallitsevat profiilit ovat kaikissa klustereissa pintasuuntautuneita sekä vertaisoppijoita näiden suuren kokonaismäärän vuoksi. Oikeanpuoleisista kuvaajista nähdään, että lähes puolet tukea tarvitsevista ja omin päin opiskelevista opiskelijoista kuuluvat ensimmäiseen klusteriin. Muuten tukea tarvitsevat ovat jakautuneet tasaisesti



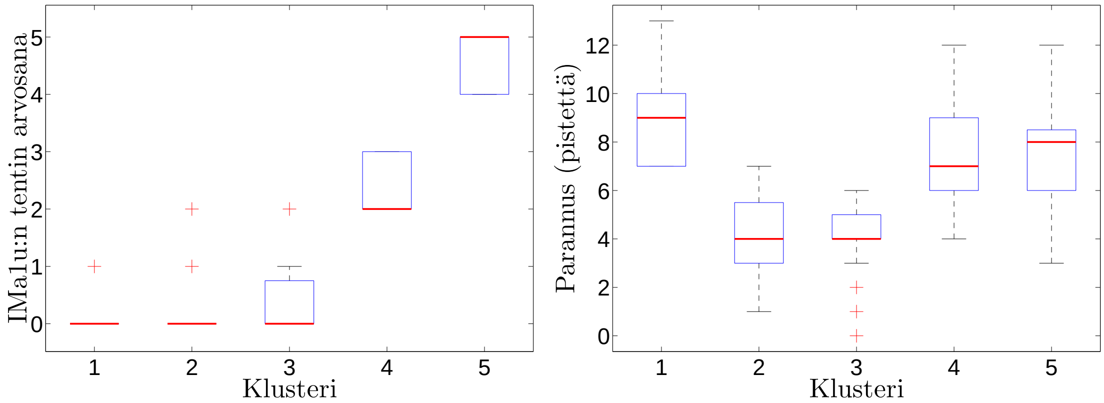
Kuva 5.18: Toisen profiilikyselyn mukaisten opiskelijaprofiiliryhmien jakautuminen klustereihin. Vasemmalla pylväät kuvaavat lukumäärää ja oikealla prosenttiosuutta opiskelijaprofiilin koosta.

toisen, kolmannen ja neljännen klusterin välille, kun tarkastellaan perustaitotestin yhteydessä pidettyä profiilikyselyä. Toisessa profiilikyselyssä profiilit ovat vaihtuneet siten, että tukea tarvitsevat ovat pääosin joko ensimmäisessä tai neljännessä klusterissa, eikä kolmannessa tai viidennessä ole yhtään tukea tarvitsevaa. Tässä vaiheessa on kuitenkin huomioitava, että kyseessä on alle kymmenen opiskelijan jakautuminen 118 opiskelijan joukossa, jolloin johtopäätösten tekeminen on riskialtista. Opiskelijaprofiilien vaihtuvuudesta klusterin sisällä havaitaan lähinnä nelosklusterin painottuminen lukumäärän suhteen vertaisoppijoiden klusteriksi. Muissa klustereissa vastaavaa yhden profiilin hallintaa ei havaita.

Paremman kuvan klustereista sekä niiden ominaisuuksista saadaan tarkastelemalla niitä tutkimuksessa käytettyjen muuttujien suhteen. Kuvassa 5.19 on esitetty vasemmalla eri klusterien laatikkokuvaajat Insinöörimatematiikka 1u:n tentin arvosanalle ja oikealla perustaitotestien väliselle parannukselle. Laatikkokuvaajien tulkinnasta kerrotaan tarkemmin aliluvussa 4.1.4 sivulla 21.

Vasemmanpuoleisesta kuvaajasta nähdään, että kolme ensimmäistä klusteria ovat tentin arvosanan perusteella tasavertaiset, ja kuten histogrammissa 5.9 sivulla 51 on esitetty, niin lähes puolet opiskelijoista sai tentistä arvosanan nolla. Tämä selittää tentissä menestymättömien klustereiden suuren määrän. Neljäs klusteri on saanut tentistä arvosanoja kaksi ja kolme, ja viides klusteri on menestynyt kaikista parhaiten saaden tentistä arvosanoja neljä ja viisi. Tämän perusteella saadaan neljännelle klusterille ominaisuudeksi keskinkertainen menestys kurssin tentissä ja viidennelle klusterille erinomainen menestys ensimmäisten jäädessä heikkolle tasolle.

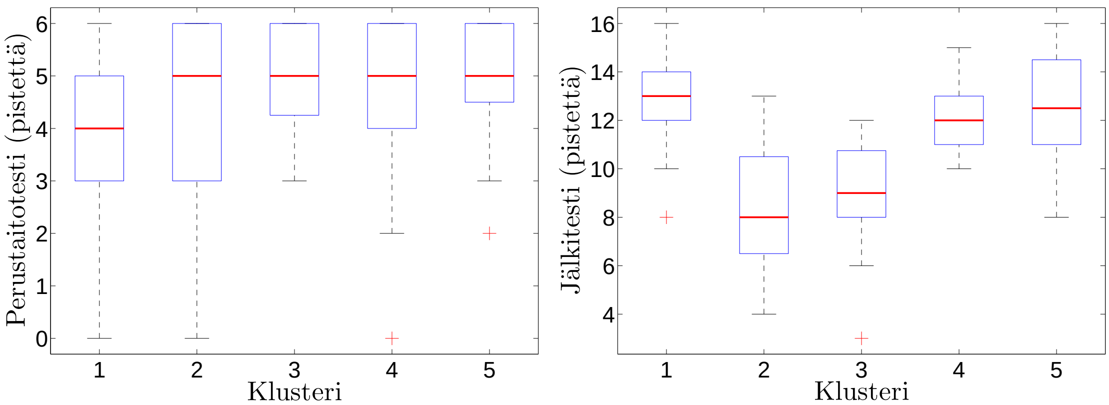
Oikeanpuoleisessa kuvaajassa saadaan hieman eroa myös ensimmäisten klustereiden välille. Ensimmäiselle klusterille on ominaista suuri parannus perustaitaitotestissä jumpan aikana. Toinen ja kolmas klusteri eivät tee suurta eroa toisiinsa myöskään parannuksen suhteen, molempien parannettua keskimäärin vain noin nel-



Kuva 5.19: Vasemmalla klusterien Insinöörimatematiikka 1u -kurssin tenttiarvosanojen laatikkokuvaajat. Oikealla klustereiden perustaitotestien välisen parannuksen laatikkokuvaajat.

jä pistettä, mikä on vähän suhteessa muihin klustereihin. Klusterit neljä ja viisi paransivat selvästi enemmän, mutta eivät päässeet ensimmäisen klusterin tasolle.

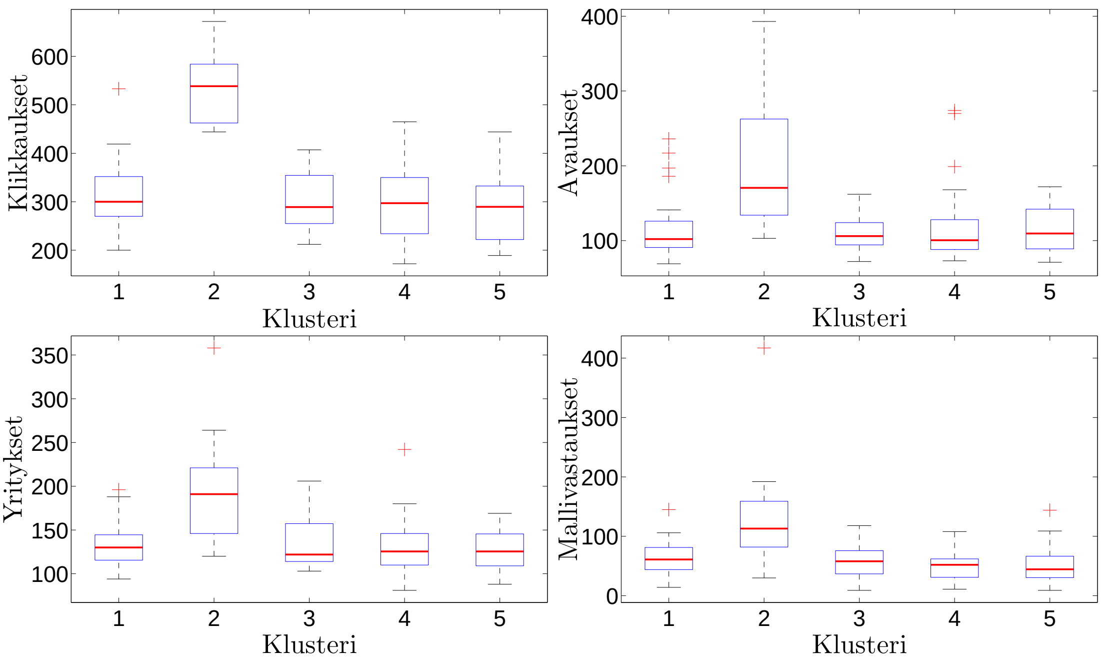
Ensimmäisen klusterin vahva parannus perustaitotesteissä mutta heikko tulos kurssin Insinöörimatematiikka 1u tentissä on selitettävissä kehityksenä Math-Bridge-tehtävien tekemisessä, eikä niinkään matematiikan taidoissa. Tällaisilla opiskelijoilla lienee ollut vaikeuksia tehdä tehtäviä tietokoneella jo perustaitotestin yhteydessä. Toisaalta opiskelijat suorittivat jälkitestin ilman valvontaa, joten apuvälineiden käyttö on ollut mahdollista, mikä selittäisi osaltaan tällaista pinnallista Math-Bridge-oppimista.



Kuva 5.20: Vasemmalla perustaitotestin tulosten laatikkokuvaajat. Oikealla vastaavat laatikkokuvaajat matematiikkajumpan jälkeisessä jälkitestissä.

Tarkastellaan tarkemmin perustaitotestin ja jälkitestin laatikkokuvaajia eri klustereissa kuvan 5.20 avulla. Kuvasta havaitaan, että parannusta kuvaavaan muuttujaan vaikuttaa eniten jälkitestin tulos. Ensimmäisen klusterin mediaani on neljä, ja muiden mediaani on viisi. Toinen klusteri on hajonnut kolmea jälkimmäistä klusteria enemmän pienempien arvosanojen suuntaan. Klustereiden välillä ei kuitenkaan ole merkittävää eroa perustaitotestin perusteella. Jälkitestissä nähdään saman suuntainen kuvaaja kuin parannuksenkin kohdalla. Eroa löytyy ainoastaan viidennen klusterin hieman suurempi hajonta neljänteen klusteriin verrattuna.

Opintomenestystä mittaavien testien suhteen eroa on löytynyt lähes kaikkien klustereiden välille. Tutkitaan seuraavaksi opiskelijoiden työmäärää matematiikkajumpassa klikkausten avulla. Klikkaukset näyttivät sivun 56 korrelaatiomatriisissa riippuvan vahvasti toisistaan. Muuttujana klusteroinnissa käytettiin klikkausten kokonaismäärää. Kuvassa 5.21 on esitetty neljä laatikkokuvaajaa, joissa on klikkausten, avauskertojen, ratkaisuyritysten sekä mallivastauksen katsomisten laatikkokuvaajat klustereittain. Kuvaajista voidaan tehdä kaksi havaintoa: ensiksikin erilaisten klikkausten jakauma klustereiden välillä on samanlainen ja toiseksi klusteri numero kaksi on tuottanut selvästi enemmän lokitapahtumia muihin klustereihin verrattuna. Toisen klusterin ominaispiirteeksi muodostuu siis suuri klikkausten määrä. Taas on huomattava, että klusterin koko on vain 12 opiskelijaa, joten suuresta ryhmästä ei siis ole kysymys. Havainto tukee kuitenkin aliluvussa 5.1.5 tehtyjä pohdintoja siitä, että klikkausten määrä heijastaa päämäärätöntä opiskelua ilman merkittäviä oppimistuloksia.



Kuva 5.21: Vasemmalla ylhäällä kokonaisklikkausmäärän laatikkokuvaajat klustereittain. Oikealla ylhäällä vastaavat laatikkokuvaajat avauksille, vasemmalla alhaalla yrityskerroille ja oikealla alhaalla mallivastausten katsomisille.

Kullakin klusterilla on siis jokin ominaispiirre, jolla se eroaa muista klustereista. Ensimmäinen klusteri oppi tekemään Math-Bridge-tehtäviä ja paransi perustaitotestin ja jälkitestin välillä selvästi, mutta oppiminen ei näkynyt peruskurssin tentissä.

Työmäärä matematiikkajumpassa oli kohtuullinen. Klusteri oli yksi suurimmista, ja tukea tarvitsevista sekä omin päin opiskelevista opiskelijoista suurin osa kuului siihen.

Toinen klusteri oli heikoimpia kaikissa oppimista mittaavissa testeissä. Ainoa merkittävä tekijä klusterin kohdalla oli valtava lokitapahtumien määrä suhteessa muihin klustereihin. Varsinkin mallivastausten katsomisten määrä oli yli kaksinkertainen muihin klustereihin verrattuna. Klusterin pienestä koosta huolimatta tämä on klusteroinnin merkittävin ryhmä, sillä näillä opiskelijoilla on hyvin todennäköisesti suuria ongelmia opiskelutavoissa.

Kolmas klusteri oli toisen klusterin kaltainen jälki- ja perustaitotestin välisessä parannuksessa sekä tentin arvosanan perusteella tehdyssä vertailussa. Ainoa merkittävä ero toiseen klusteriin oli kuitenkin vähäinen työmäärä matematiikkajumpassa. Klusteri oli tutkimuksen suurin, ja siihen kuului yli neljännes kaikista jumppaajista. Klusterin sisällä opiskelijat vaihtoivat opiskelijaprofiilia niin, että kukaan ei kokenut olevansa tukea tarvitseva ja yhä useampi oli vertaisoppija tai osaaja. Lukumäärällisesti suurin profiili oli kuitenkin pintasuuntautuneet mallista oppijat.

Neljännellä klusterilla Insinöörimatematiikka 1u:n tentin arvosana oli jo selvästi hyväksytyn puolella, ja perustaitotestissäkin opiskelijat paransivat hyvin. Työmäärä ei ollut merkittävästi erilainen kuin muilla klustereilla toista klusteria lukuun ottamatta. Klusteri oli kooltaan keskitasoa, ja sitä hallitsivat opiskelijaprofiileista vertaisoppijat.

Viides klusteri menestyi erinomaisesti kurssin tentissä ja parannusta tuli perustaitotestin ja jälkitestin välillä paljon. Klusteri oli pienehkö, ja se sai merkittäviä oppimistuloksia aikaiseksi pienellä määrällä klikkauksia. Tällaisilla opiskelijoilla voidaan todeta olevan hyvät pohjatiedot matematiikassa tai ainakin hyvä motivaatio ja tapa opiskella ja kerrata matematiikan asioita. Todennäköisesti opiskelijoiden suorittama perusasioiden kertaus on osoittautunut hyödylliseksi ja tämä on edesauttanut oppimista myös peruskurssilla pienentäen riskiä pudota kurssin vaatimasta opiskelutahdista.

Opiskelijat voidaan siis jakaa klusterianalyysin perusteella viiteen ryhmään:

Jumppaajat: Opiskelija oppii jumpan aikana tekemään pinnallisesti Math-Bridge-tehtäviä, mutta syvempi matematiikan oppiminen jää saavuttamatta.

Häslääjät: Opiskelijalla on suuria vaikeuksia matematiikan opiskelussa. Tämä näkyy matematiikkajumpassa päämäärättömänä klikkailuna ja testeissä heikkoina tuloksina.

Tuntemattomat: Opiskelija ei menestynyt testeissä, eikä hänen toiminnastaan voida jumpan perusteella tehdä mitään erityisiä johtopäätöksiä.

Keskikasti: Opiskelija hyötyi jumpasta ja pääsi hyvään vauhtiin matematiikan opiskelussa. Tuloksellisesti suoritus on normaalin opiskelijan tasolla.

Hyötyjät: Opiskelija sai jumpasta täyden hyödyn osaamansa asian kertaamiseen, joka näkyi hyvinä testi- ja tenttituloksina.

5.4 Ehdollinen todennäköisyystaulu

Aikaisempien tutkimusten kohdalla aliluvussa 2.3 sivulla 5 esiteltiin Arroyon ja Woolfin käyttämä ehdolliseen todennäköisyyteen perustuva taulukko, jossa aineistoa jaetaan mediaanien perusteella puoliksi haluttujen muuttujien avulla. Näin opiskelijat saadaan ryhmiteltyä ja laskettua todennäköisyyksiä kuhunkin ryhmään kuulumiselle eri muuttujien vaikutuksesta. Arroyo ja Woolf rakensivat aluksi Bayesin verkon, jonka perusteella he valitsivat sopivat muuttujat taulukkoon. [2]

Tässä tutkimuksessa opiskelijat jaettiin ryhmiin kolmen muuttujan, klikkausten kokonaismäärän, Insinöörimatematiikan tenttiarvosanan sekä jälki- ja perustaitotestin välisen parannuksen perusteella. Muuttujat valittiin ilman tilastollisia menetelmiä. Ne kuvaavat kuitenkin parhaiten niitä asioita, joita haluttiin tutkia.

Taulukko 5.12: Muuttujien mediaanit sekä opiskelijoiden jakautuminen mediaania pienempiin sekä suurempiin tai yhtä suuriin.

Muuttuja	Mediaani	Opiskelijoiden lukumäärä ($<$ / $\geq$)
Klikkaukset	303,5	59/59
IMA1u	1	56/62
Parannus	7	57/61

Muuttujille laskettiin mediaani, jonka perusteella opiskelijat pyrittiin jakamaan kahteen osaan, jotka ovat mahdollisimman lähelle saman kokoiset. Taulukossa 5.12 on esitetty muuttujien mediaanit ja opiskelijoiden lukumäärät mediaanin alapuolella sekä mediaanissa ja sen yläpuolella. Mediaani jakaa opiskelijat klikkausten määrän suhteen sopivasti kahteen yhtä suureen osaan. Insinöörimatematiikan tentin kohdalla valinta tehtiin siten, että ryhmät ovat "Hylätty tenttitulos" ja "Hyväksytty tenttitulos". Tämä on hyvä ja intuitiivinen jakoperuste. Parannuksen kohdalla raja vedettin mediaanin alapuolelle, jolloin opiskelijat saatiin puolitettua melko hyvin.

Taulukossa 5.13 on esitetty ehdollisen todennäköisyyden taulukko. Taulukkoa voidaan lukea monella eri tavalla ja luoda erilaisia todennäköisyyksiä ryhmien lukumäärien perusteella. Taulukon avulla saadaan vahvistettua kuitenkin niitä havaintoja, joita tehtiin korrelaatioiden ja klusteroinnin yhteydessä. Vähän klikanneista (ryhmät 1–4) tentin suoritti hyväksytysti noin 57,3 % ja paljon klikanneista (ryhmät 5–8) tentissä hylättiin 52,5 %. Lisäksi vähän klikanneista 59,3 % paransi perustaitotestin ja jälkitestin välillä paljon, kun paljon klikanneista 55,9 % opiskelijaa paransi testeissä vain vähän.

Taulukko 5.13: Ehdollisen todennäköisyyden taulukko.

	Puolittavat muuttujat							
Klik.	303 tai alle				yli 303			
IMA1u	Hylätty		Hyväksytty		Hylätty		Hyväksytty	
Parannus	0–6	7–16	0–6	7–16	0–6	7–16	0–6	7–16
Ryhmä#	1	2	3	4	5	6	7	8
	Muuttujien keskiarvot							
Parannus	4,00	8,92	4,33	8,45	3,90	9,09	5,46	8,73
PTT	5,25	4,15	5,42	4,45	4,60	3,82	4,92	4,40
JäT	9,25	13,08	9,75	12,91	8,50	12,91	10,38	13,13
IMA1u	0,00	0,00	2,25	3,09	0,00	0,00	2,46	2,80
Klik.	261,25	271,00	259,58	237,55	437,40	389,27	402,15	365,40
Avaukset	102,25	97,46	97,50	92,73	139,45	146,36	168,77	159,13
Malliv.	40,17	44,31	42,00	35,14	114,05	92,73	78,38	63,80
Yritykset	118,83	129,23	120,08	109,68	181,90	150,18	155,00	142,47
Oikein	72,67	72,85	73,58	72,55	77,70	81,82	75,92	75,93
Väärin	46,17	56,38	46,50	37,14	104,20	68,36	79,08	66,53
	Ryhmien lukumäärät							
Opiskeli-	12	13	12	22	20	11	13	14
joiden lu-	25		34		31		28	
kumäärä	59				59			

Merkittävää on myös se, että vähän klikanneista ja hyväksytyn arvosanan saaneista opiskelijoista (ryhmät 3 ja 4) suurin osa (64,7 %) paransi paljon testien välillä. Toisaalta paljon klikanneet ja hylätyn arvosanan tentistä saaneet opiskelijat (ryhmät 5 ja 6) olivat todennäköisemmin (64,5 %) myös parantaneet vain vähän testien välillä. Tarkastelemalla muuttujien keskiarvoja ryhmien välillä havaitaan, että ryhmät 4 ja 5 ovat mielenkiintoisia. Ryhmällä 4 on korkein tentin keskiarvo ja matalin keskiarvo kaikissa klikkauksien määrää kuvaavissa muuttujissa. Toisaalta ryhmällä 5 on suurimmat keskiarvot lähes kaikissa klikkamista kuvaavissa muuttujissa ja parannus testien välillä on kaikista ryhmistä pienintä.

6. YHTEENVETO

Tutkimukselle asetettiin luvussa 3.2 viisi tutkimuskysymystä. Tässä luvussa vedetään yhteen tutkimuksen eri osat vastaamalla tutkimuskysymyksiin tutkimuksen pohjalta.

1. Käyttäytyvätkö erilaiset opiskelijaprofiilit eri tavoin matematiikkajumpassa? Jumpan lokitiedoista kerättyihin muuttujiin verrattessa havaittiin lähinnä pinta- ja vertaisoppijoiden keskimääräistä heikompi suoritustaso testien välisessä parannuksessa sekä suurehko klikkausten määrä. Lisäksi havaittiin tukea tarvitsevien heikko menestys Insinöörimatematiikan ensimmäisessä tentissä sekä omin päin opiskelevien verrattain pieni klikkausten määrä. Osaajien kannalta tarkastelu muuttujien suhteen paljasti joidenkin opiskelijoiden kohdalla virheellisiä käsityksiä omasta opiskelusta. Vaihtuvuus profiilien välillä kuitenkin kertoi opiskelijoiden kehittyneestä kyvystä oman opiskelunsa analysoinnissa.

2. Voidaanko lokitietojen perusteella ryhmitellä opiskelijoita? Lokitiedoista saatuja muuttujia verrattessa erilaisista testeistä saatuihin muuttujiin, löydettiin yhteys, jonka perusteella klikkausten määrällä mitattu aktiivisuus korreloisi negatiivisesti oppimistulosten kanssa. Tästä havainnosta saatiin hyvä lähtökohta opiskelijoiden toiminnan tarkemmalle tutkimiselle ja sen pohjalta saatiin valittua sopivat muuttujat opiskelijoiden ryhmittelyä varten. Lisäksi kyseistä havaintoa vahvistettiin varsinaisen klusterianalyysin lisäksi myös ehdollisen todennäköisyyden taulukolla. Tämä tulos näyttää olevan ristiriidassa Arroyon ja Woolfin tutkimuksen [2] kanssa.

3. Voidaanko kohdan 2 ryhmien perusteella löytää opiskelijat, joilla on ongelmia matematiikan opiskelussa? Opiskelijoiden ryhmittely testi- ja lokitietojen avulla antoi mahdollisuuden tutkia toisaalta niitä opiskelijoita, jotka ovat vaarassa ajautua ongelmiin matematiikan opiskelussa jo yliopiston peruskursseilla. Näiden opiskelijoiden tunnistamiseen voidaan hyödyntää matematiikkajumpan lokitietoja, mutta matematiikkajumpasta ei tällaisenaan ole näiden opiskelijoiden pelastajaksi.

4. Kenelle matematiikkajumppa on hyödyllinen? Matematiikan opiskelussa mahdollisesti ongelmiin ajautuvien opiskelijoiden lisäksi tutkimuksessa havaittiin myös ne ryhmät, joille matematiikkajumpasta oli hyötyä. Näillä opiskelijoilla jumppa toimi hyvänä keinona kerrata matematiikan osa-alueita, jotka ovat jääneet aikaisemmissa opinnoissa vähemmälle tai ajan myötä unohtuneet. Näissä opiskelijoissa oli mukana sellaisia, jotka olisivat todennäköisesti jumpasta huolimatta kerranneet opiskeltavat asiat, mutta myös niitä, jotka selvästi pääsivät muiden opiskelijoiden tasolle matematiikkajumpan avulla.

5. Voidaanko lokitietojen perusteella havaita ongelmallisia tehtäviä ja tehtävätyyppejä? Matematiikkajumpasta havaittiin myös tehtäviä, jotka olivat aiheuttaneet paljon klikkauksia. Osa näistä tehtävistä eivät olleet mielekkäitä jumpan tarkoituksen kannalta, vaan aiheuttivat ongelmia muiden kuin ratkaisutaitojen vuoksi. Tehtävien pituudessa oli paljon eroa, jolloin pienet huolimattomuusvirheet kostautuivat tiettyjen tehtävien kohdalla näkyen lisääntyneinä ratkaisuyrityksinä. Toisaalta tehtävistä nähtiin opiskelijoille lähtökohtaisesti vaikeimmat osa-alueet, jotka olivat integrointi, derivointi ja epäyhtälöiden ratkaiseminen.

Tutkimuksessa käytetyt lokitiedot antoivat hyvät mahdollisuudet opiskelijoiden tutkimiseen, mutta niistä ei saatu käytettyä kaikkea mahdollista potentiaalia. Lokiin tallentuneista aikatiedoista voisi jatkossa saada lisää hyviä opiskelijoiden opiskelua kuvaavia muuttujia. Lisäksi jatkotutkimuksissa voisi tutkia tarkemmin erilaisten klikkausten kautta opiskelijoiden opiskelutapoja. Tehtävien kannalta olisi mielenkiintoista tutkia jatkossa myös tehtävien ratkaisupolkuja järjestyksen ja ajan suhteen sekä verrata niitä opiskelijoiden kesken.

Mahdollisten parannus- ja jatkotutkimusten tarpeesta huolimatta tutkielmassa esitetyt tutkimustulokset antavat vastauksia alussa asetettuihin tutkimuskysymyksiin. Näin ollen tutkimusta voidaan pitää onnistuneena jatkona aikaisempiin Tampereen teknillisen yliopiston matematiikan laitoksen opetuksen kehittämistä koskeviin tutkimuksiin.

LÄHTEET

[1] Alexander, S. AusWeb95 The First Australian WorldWideWeb Conference, Ballina, Australia, 30.4.–2.5. 1995. [WWW]. [viitattu 12.2.2013]. Saatavissa: `http://ausweb.scu.edu.au/aw95/education2/alexander/`.

[2] Arroyo, I. & Woolf, B.P. *Inferring learning and attitudes from a Bayesian Network of log file data*. Artificial Intelligence in Education, Amsterdam, 2005. Amsterdam 2005, IOS Press. pp. 33–40.

[3] Biehler, R., Fischer, P.R., Hochmuth, R., Jeuring, J. & Wassong, T. *How to support students learning in mathematical bridging-courses using an ITS? Remedial scenarios in the European project Math-Bridge*. Utrecht University 2010, Department of Information and Computing Sciences, Tekninen raportti UU-CS-2010-023. 10 p.

[4] Cocea, M. & Weibelzahl, S. *Can log files analysis estimate learners' level of motivation?* 14th Workshop on Adaptivity and User Modeling in Interactive Systems, Hildesheim, Germany, October 9–11, 2006. Hildesheim 2006. pp. 32–35.

[5] European Comission, Europe's Information Society. *Math-Bridge: European Remedial Content for Mathematics* [WWW]. [viitattu 31.10.2012]. Saatavissa: `http://ec.europa.eu/information_society/apps/projects/factsheet/index.cfm?project_ref=ECP-2008-EDU-428046`.

[6] Gorard, S. *Revisiting a 90-year-old debate: the advantages of the mean deviation.* British Journal of Education 53(2005)4, pp. 417–430.

[7] Jain, A.K., Murty, M.N. & Flynn, P.J. *Data Clustering: A Review*. ACM Computing Surveys 31(1999)3, pp. 264–323.

[8] Johnson, R.A. & Wichern, D.W., *Applied multivariate statistical analysis*. New Jersey 1992, Prentice-Hall, Inc. 642 p.

[9] Kriegel, H.P., Kröger, P., Sander, J. & Zimek, A. *Density-based clustering*. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery 1(2011)3. pp. 231–240.

[10] Lee, Y.-J. *Developing an efficient computational method that stimates the ability of students in a Web-based learning environment*. Computers & Education 58 (2012), pp. 579-589.

[11] Lilliefors H.W. *On the Kolmogorov–Smirnov Test for Normality with Mean and Variance Unknown*. Journal of the American Statistical Association 62(1967)318, pp. 399–402.

[12] Massey, F.J. *The Komogorov-Smirnov Test for Goodness of Fit*. Journal of the American Statistical Association 46(1951)253, pp. 68–78.

[13] Metsämuuronen, J. *Tutkimuksen tekemisen perusteet ihmistieteissä*. 2. painos. Gummerus Kirjapaino Oy, Jyväskylä 2003. 772 s.

[14] Milton, J.S. & Arnold, J.C. *Probability and Statistics in the Engineering and Computing Sciences*. Singapore 1986, McGraw-Hill Book Co. 643 p.

[15] Pohjolainen, S., Raassina, H., Silius, K., Huikkola, M., & Turunen, E. *TTY:n insinöörmatematiikan opiskeljoiden asenteet, taidot ja opetuksen kehittäminen*. Tampere 2006, Tampereen teknillinen yliopisto, Matematiikan laitos, Tutkimusraportti 84. 80 s. + liit. 43 s.

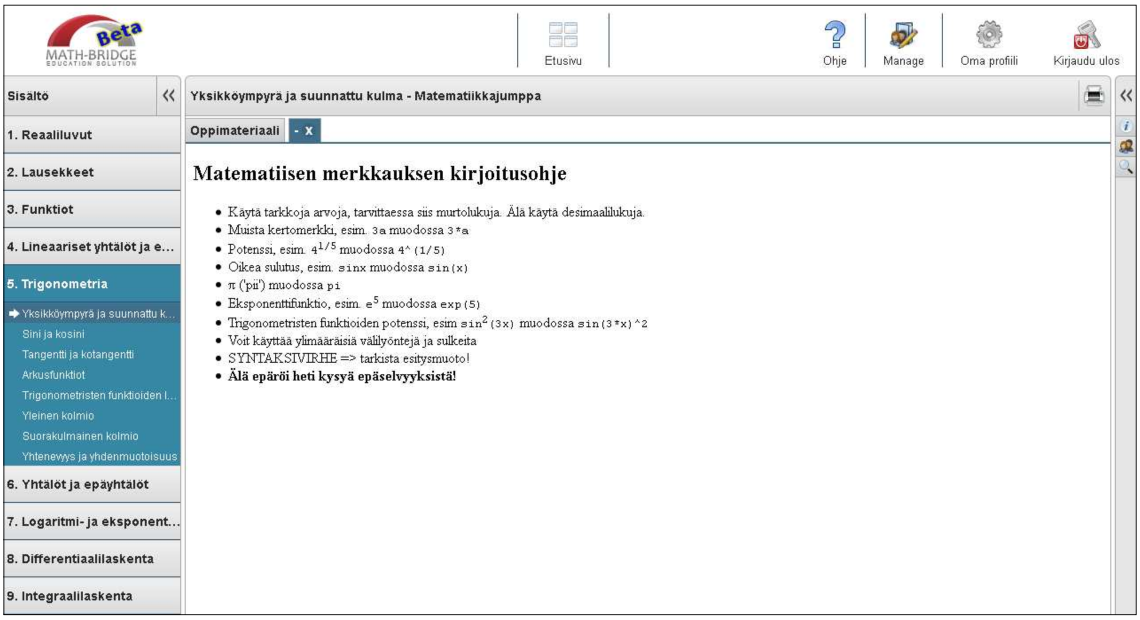
[16] VanLehn, K., Lynch, C., Schulze, K., Shapiro, J.A., Shelby, R., Taylor, L., Treacy, D., Weinstein, A. & Wintersgill, M. *The Andes Physics Tutoring System: Lessons Learned*. International Journal of Artificial Intelligence in Education 15(2005)3, pp. 147–204.

[17] Walpole, R.E., Myers, R.H., Myers, S.L. & Ye, K. *Probability & Statistics for Engineers & Scientists*. 8th ed. Pearson Prentice Hall 2007. 816 p.

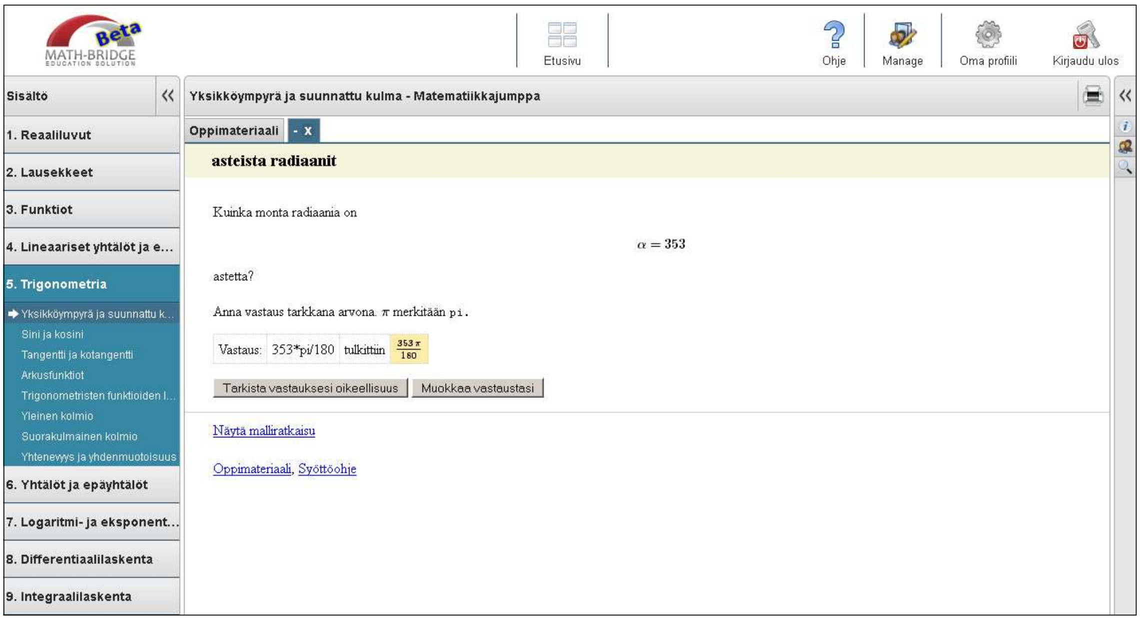
[18] Xu, X., Ester, M., Kriegel, H.-P. & Sander, J. *A distribution based clustering algorithm for mining in large spatial databases*. 14th International Conference on Data Engineering, Orlando, Florida, USA, February 23–27, 1998. pp. 324–331.

A. MATH-BRIDGE

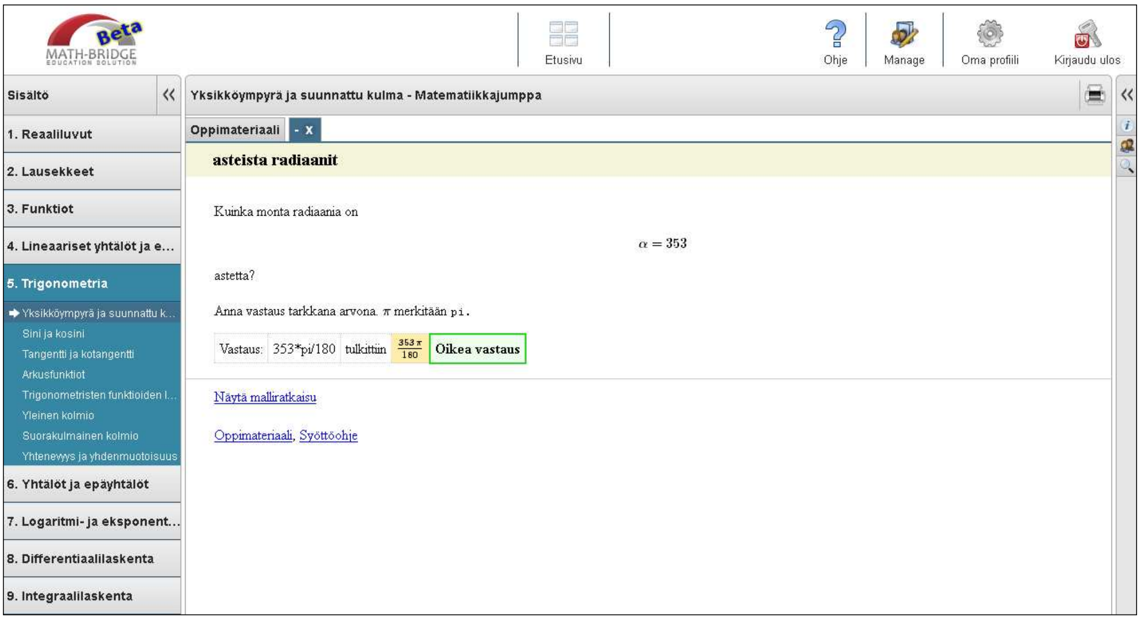
Tässä liitteessä esitellään kuvakaappauksia Math-Bridge-ohjelmasta. Muita Math-Bridgen käyttäjälle avautuvia näkymiä on esitelty aliluvussa 2.2.1 sivulla 3.



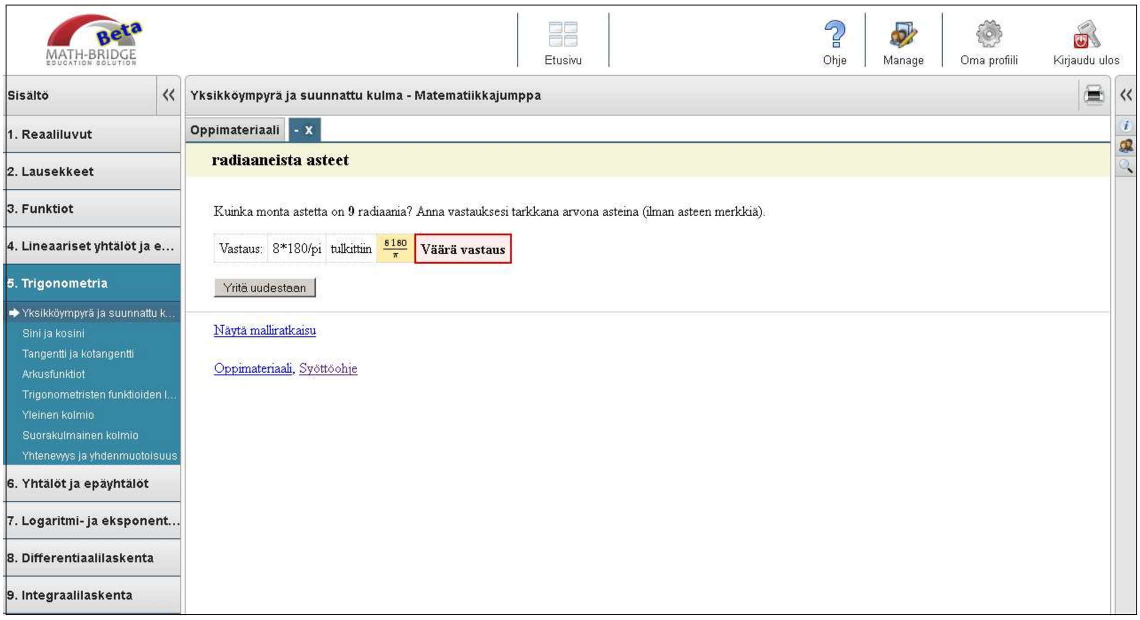
Kuva A.1: Math-Bridgen näkymä käyttäjän tarkastellessa matemaattisten merkkien kirjoitusohjetta.



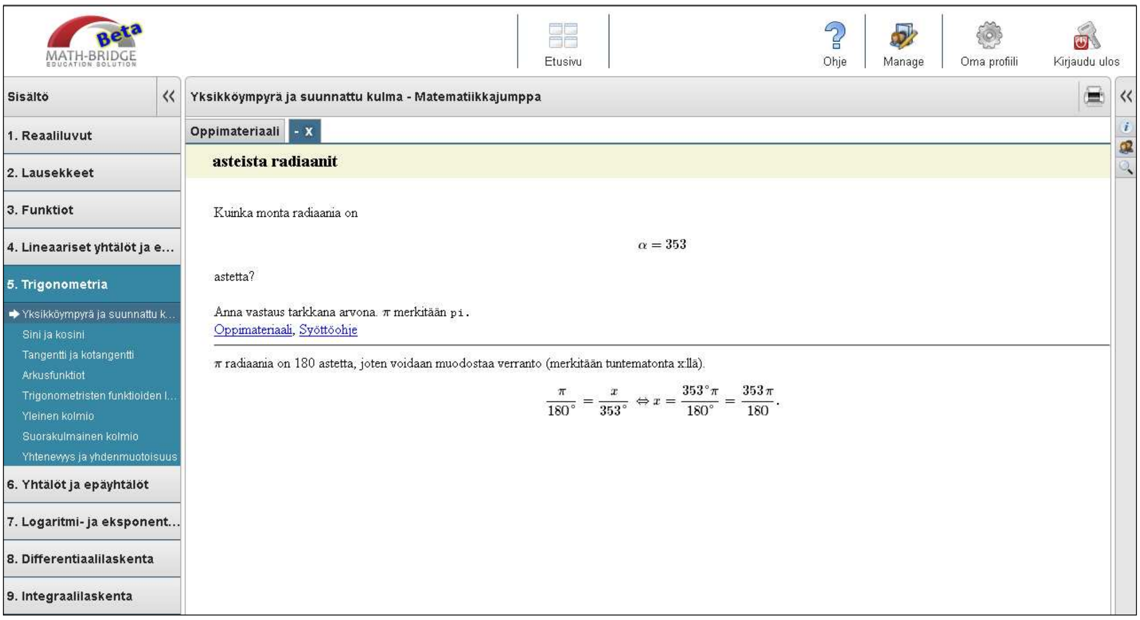
Kuva A.2: Math-Bridgen näkymä kun käyttäjä on tarkistanut syötetyn ratkaisun syntaksin. Käyttäjä voi tässä vaiheessa korjata syntaksia tarpeen mukaan.



Kuva A.3: Math-Bridgen näkymä kun käyttäjä on syöttänyt oikean ratkaisun ja tarkistanut vastauksen oikeellisuuden.



Kuva A.4: Math-Bridgen näkymä kun käyttäjä on syöttänyt väärän ratkaisun. Käyttäjä voi tämän jälkeen tarkastella malliratkaisua tai yrittää uudelleen.



Kuva A.5: Math-Bridgen näkymä kun käyttäjä tarkastelee tehtävän malliratkaisua.

B. LOKITIEDOT JA DATAMATRIISI

Tässä liitteessä esitellään esimerkkitaulukot lokitiedoista ja datamatriisista. Malliksi on otettu vain pieni osa kustakin aineistosta havainnollistamaan tiedostojen muotoa. Lokitiedoista on muutettu opiskelijanumero kirjaimeksi A ja datamatriisissa opiskelijat on numeroitu juoksevalla numeroinnilla. Lokitiedot ovat siinä muodossa kuin ne olivat ennen Python-ohjelmalla parsimista.

Taulukko B.1: Esimerkki lokitiedoista. Lokitietojen sarakkeet on esitelty tarkemmin aliluvussa 3.4 sivulla 12.

Aikaleima	Op.num.	Tapahtuma	Teht. koodi	Palaute
19.9.2011 14:05	A	ExerciseStarted	der02	
19.9.2011 14:06	A	ExerciseStep	der02	1
19.9.2011 14:07	A	ExerciseStarted	der02	
20.9.2011 22:19	A	ExerciseStarted	der02	
20.9.2011 22:25	A	ExerciseStarted	der02	
20.9.2011 22:26	A	ExerciseStep	der02	1
21.9.2011 15:57	A	ExerciseStarted	n1	
21.9.2011 15:58	A	ExerciseStep	n1	1
21.9.2011 15:58	A	ExerciseStarted	n2	
21.9.2011 15:59	A	ExerciseStep	n2	1
21.9.2011 15:59	A	ExerciseStarted	n3	
21.9.2011 15:59	A	ExerciseStep	n3	1
21.9.2011 15:59	A	ExerciseStarted	n4	
21.9.2011 15:59	A	ExerciseStep	n4	1
21.9.2011 15:59	A	ExerciseStarted	n5	
21.9.2011 16:00	A	ExerciseStep	n5	1
21.9.2011 16:02	A	ExerciseStarted	n8	
21.9.2011 16:02	A	ExerciseStep	n8	0
21.9.2011 16:03	A	ExerciseStep	n8	0
21.9.2011 16:03	A	ExerciseStep	n8	0
21.9.2011 16:04	A	ExerciseStep	n8	0
21.9.2011 16:04	A	ExerciseStep	n8	-1
21.9.2011 16:05	A	ExerciseStarted	n8	
21.9.2011 16:06	A	ExerciseStep	n8	1
21.9.2011 16:06	A	ExerciseStarted	n5	
21.9.2011 16:06	A	ExerciseStep	n5	1
21.9.2011 16:07	A	ExerciseStarted	n4	

Taulukko B.2: Esimerkki datamatriisista. Matriisi on katkaistu tehtäväkohtaisten tietojen kohdalta. Matriisin sarakkeet ovat opiskelijanumero (O#), tehtäväkohtainen data (viisi saraketta), katkaisumerkki (···), koulutusohjelma (ko), perustaitotesti (ptt), opiskelijaprofiili 1 (p1), jälkitesti (jät), opiskelijaprofiili 2 (p2), Insinöörimatematiikka 1u:n tenttiarvosana (IMa), sukupuoli (s) sekä avausten (ava), yritysten (yri) ja mallivastausten (mv) kokonaismäärät. Datamatriisin sarakkeet on esitelty tarkemmin aliluvussa 3.5 sivulla 15.

O#	der2 (avaus, yritys, malliv, avattu, ratkaistu)					···	ko	ptt	p1	jät	p2	IMa	s	ava	yri	mv
1	8	6	0	19.9.2011 14:05	19.9.2011 14:06	···	1161	6	4	9	5	0	2	124	158	100
2	4	1	1	21.9.2011 19:53	21.9.2011 19:55	···	1051	6	1	12	2	5	1	149	116	67
3	1	1	0	22.9.2011 20:44	22.9.2011 20:47	···	1046	3	3	8	5	0	1	135	184	75
4	5	1	2	20.9.2011 16:37	20.9.2011 18:33	···	1137	4	1	12	4	0	2	116	114	45
5	5	1	0	19.9.2011 11:26	19.9.2011 11:27	···	1137	5	4	11	4	0	1	89	112	26
6	4	2	3	19.9.2011 19:01	19.9.2011 19:08	···	1144	0	1	4	1	0	1	103	151	417
7	2	1	0	19.9.2011 18:12	7.10.2011 8:33	···	1231	5	4	10	4	1	1	93	110	15
8	6	4	0	21.9.2011 15:53	21.9.2011 15:54	···	1051	5	2	9	1	0	1	130	133	63
9	2	2	0	24.9.2011 11:11	24.9.2011 11:12	···	1137	4	2	11	2	2	1	96	121	62
10	2	1	0	26.9.2011 17:13	26.9.2011 18:04	···	1052	2	3	12	1	0	1	100	196	106
11	1	1	0	21.9.2011 21:32	21.9.2011 21:33	···	1051	3	2	15	2	5	1	135	125	73
12	1	1	0	26.9.2011 20:58	26.9.2011 20:59	···	1052	5	1	13	1	1	1	69	102	29
13	6	6	0	26.9.2011 11:44	26.9.2011 11:45	···	1046	5	3	10	2	2	1	89	176	102
14	4	2	1	8.10.2011 16:47	8.10.2011 17:01	···	1052	6	1	10	1	0	1	135	113	36
15	6	3	1	19.9.2011 12:02	19.9.2011 12:11	···	1137	3	5	11	5	3	1	80	81	11
16	1	1	0	30.9.2011 19:22	30.9.2011 19:23	···	1137	6	1	13	1	2	1	73	89	13
17	3	1	0	20.9.2011 17:32	20.9.2011 17:32	···	1231	5	1	12	1	0	1	95	130	61
18	1	1	0	28.9.2011 11:14	28.9.2011 11:14	···	1231	5	4	11	1	0	1	118	135	100
19	1	1	0	19.9.2011 18:59	19.9.2011 18:59	···	1225	6	1	12	2	1	1	143	206	58
20	1	1	0	25.9.2011 20:52	25.9.2011 20:53	···	1052	5	1	13	1	4	1	71	88	30

C. TEHTÄVÄTYYPIT

Tässä liitteessä esitellään matematiikkajumpan tehtävät aihealueittain sekä erilaisten tapahtuminen jakautuminen tehtävittäin. Opiskelijoille annettiin tehtävänannossa ohjeistusta myös tehtävässä käytettävistä merkinnöistä, kuten neliöjuuresta ($\sqrt{(x)} = \text{sqrt(x)}$), trigonometristen funktioiden potensseista ($\sin^6(x) = \text{sin(x)\^{}6}$) sekä kreikkalaisista aakkosista ($\pi = \text{pi}$). Math-Bridge generoi tehtävän numerovakiot uudelleen joka kerta kun tehtävä avataan.

C.1 Derivointi

der2: Laske funktion

$$f(x) = 3x^5 + 4$$

derivaatta.

der4: Laske funktion

$$f(x) = (8 - 8x)^4$$

derivaatta.

der5: Laske funktion

$$f(x) = x^3 - 75x$$

pienin arvo välillä [0,5].

der6: Laske funktion

$$f(x) = 6\cos(9x) + 4\sin(8x) + 5\sin^3(4x) + 2\cos^5(3x)$$

derivaatta.

der7: Laske funktion

$$f(x) = \ln(5x) + 3e^{2x}$$

derivaatta.

der8: Derivoi funktio

$$f(x) = \ln\left(\frac{x-2}{x^2-4}\right).$$

der9: Derivoi funktio

$$f(x) = \frac{1}{x^3}.$$

der10: Derivoi funktio

$$f(x) = \left(9x^2 + 7\right)^{\frac{1}{4}}.$$

der11: Derivoi funktio

$$f(x) = e^{-\frac{9x}{2}}\left(\ln(4x)\right)^{\frac{1}{9}}.$$

C.2 Yhtälöt

eq1: Ratkaise 1. asteen yhtälö

$$10 - 9x = 0$$

Anna vastaus tarkkana arvona murtolukumuodossa.

eq2: Laske 2. asteen yhtälön

$$8x - 3x^2 = 3$$

diskriminantin arvo.

eq3: Ratkaise 3. asteen yhtälö

$$2x^3 - 20x = 0.$$

Vastaukseksi tulee kolme x:n arvoa. Anna vastaus tarkkoina arvoina.

eq4: Määritä itseisarvoyhtälön

$$|5x + 2| = 5 - x$$

ratkaisut. Vastaukseksi tulee kaksi x:n arvoa. Anna vastaus tarkkoina arvoina.

eq5: Ratkaise yhtälö

$$\sqrt{4x + 4} = 3.$$

Anna vastaus tarkkana arvona.

eq6: Ratkaise murtoyhtälö

$$\frac{x^2}{x - 6} + 3 = 0.$$

Vastaukseksi tulee kaksi x:n arvoa. Anna vastaus tarkkoina arvoina.

eq7: Ratkaise murtoyhtälö

$$\frac{9}{x - 2} - \frac{18}{x^2 - 4} = 3.$$

Vastaukseksi tulee kaksi x:n arvoa. Anna vastaus tarkkoina arvoina.

C.3 Lausekkeet

exp1: Laske funktion

$$f(x) = \sqrt{x} + \frac{1}{x} - 1$$

arvo, kun x saa arvon $x = 9$. Anna vastaus tarkkana arvona.

exp3: Sievennä lauseke

$$\frac{a^2 - 64}{a - 8}$$

(Olettaen, että nimittäjä ei ole 0).

exp4: On annettu funktio

$$f(x) = x^2 + 1.$$

Laske yhdistetyn funktion $(f \circ f)(x)$ arvo pisteessä $x = 2$.

exp5: Laske funktion

$$f(x) = 2x + 8$$

käänteisfunktio. Anna ratkaisemasi funktio muuttujan x avulla esitettynä.

exp6: Sievennä lauseke

$$\frac{1}{9x + 1} + \frac{8}{x + 1}$$

murtolausekkeeksi. Anna saamasi lauseke kokonaisuudessaan.

exp7: Määritä vakioiden A, B, C ja D arvot siten, että lauseke

$$4x + \frac{4}{x + 5} - \frac{3}{x^2 - 25} = \frac{Ax^3 + Bx^2 + Cx + D}{x^2 - 25}$$

on tosi. Anna vakioiden A, B, C ja D arvot tarkkoina arvoina.

C.4 Epäyhtälöt

ieq2: Ratkaise ensimmäisen asteen epäyhtälö

$$10x + 4 < 0.$$

Anna vastauksesi muodossa $x < a$ tai $x > a$, missä a on juuren tarkka arvo.

ieq3: Ratkaise 2. asteen epäyhtälö

$$5x^2 + 5x - 10 > 0.$$

Epäyhtälö toteutuu kaikilla x:n arvoilla arvosta a lähtien. Anna vastaukseksi kyseinen a:n arvo.

ieq4: Funktio f on aina positiivinen eräällä välillä $(-\infty, c)$.

$$f(x) = 5x^2 - 5x - 10$$

Mikä on c:n arvo?

ieq5: Ratkaise itseiarvoepäyhtälö

$$|3 - 4x| < 5.$$

Epäyhtälö on voimassa eräällä välillä (a, b). Määrää välin alaraja a. Anna vastaus tarkkana arvona.

ieq6: Ratkaise epäyhtälö

$$\frac{4x}{3} < \frac{1 - 6x}{x - 8}.$$

Vastaus on muotoa $x < a$ tai $b < x < c$. Anna luvut a, b ja c tarkkoina arvoina.

ieq7: Ratkaise epäyhtälö

$$\left| \frac{3x + 1}{7x - 1} - \frac{3}{7} \right| < 21.$$

Vastaus on muotoa $x < a$ tai $x > b$. Anna luvut a ja b tarkkoina arvoina.

ieq8: Ratkaise itseiarvoepäyhtälö

$$\left| \frac{x + 5}{x - 1} \right| < 4.$$

Vastaus on muotoa $x < a$ tai $x > b$. Anna luvut a ja b tarkkoina arvoina.

C.5 Integrointi

int1: Olkoon annettu funktio

$$f(x) = -2x - 9.$$

Laske funktion integraalifunktio (ilman integrointivakiota).

int2: Määritä funktion

$$f(x) = 10 - 7x^2$$

integraalifunktio (ilman integrointivakiota).

int3: Määritä funktion

$$f(x) = \frac{1}{(x-1)^6}$$

jokin integraalifunktio (ilman integrointivakiota).

int4: Laske määrätty integraali

$$\int_{-\pi/2}^{\pi/2} 4\cos x\sqrt{16 - 16\sin^2 x}\, dx.$$

Huom! Anna vastaus tarkkana arvona.

int5: Laske määrätty integraali

$$\int_{-1}^{3} e^{-5x}\, dx.$$

Anna vastaus tarkkana arvona. e merkitään e.

int6: Laske määrätty integraali

$$\int_{-1}^{4} 5x^4 + 4x^2 - 4\, dx.$$

Anna vastaus tarkkana arvona.

int7: Laske määrätty integraali

$$\int_{-2}^{2} -5\left|x^2 + x\right|\, dx.$$

Anna vastaus tarkkana arvona.

int8: Laske määrätty integraali

$$\int_{0}^{4} 2e^{x}\, dx.$$

C.6 Raja-arvot

lim1: Määritä seuraava raja-arvo

$$\lim_{x\to 8} x^2 + 8x + 1.$$

Anna vastaus tarkkana arvona.

lim2: Määritä seuraava raja-arvo

$$\lim_{x\to 10} \frac{x^2 - 100}{x - 10}.$$

Anna vastaus tarkkana arvona.

lim3: Määritä seuraava raja-arvo

$$\lim_{x\to 11} \frac{x^2 - 17x + 66}{x - 11}.$$

Anna vastaus tarkkana arvona.

lim4: Määritä seuraava raja-arvo

$$\lim_{x\to 0} \frac{\sin^2(7x)}{1 - \cos(7x)}.$$

lim5: Määritä b siten, että seuraava yhtälö toteutuu:

$$2 = \lim_{x\to 3} 6x^3 - 3x^2 + b.$$

Anna vastaus tarkkana arvona.

lim6: Määritä vakio b siten, että seuraava funktio on kaikkialla jatkuva. Funktio f on määritelty paloittain seuraavasti:

$$f(x) = \begin{cases} bx - 10, \text{ kun } x < 3 \\ x^2 - 3, \text{ kun } x \geq 3. \end{cases}$$

Anna vastaus tarkkana arvona.

C.7 Logaritmit

log1: Ratkaise logaritmiyhtälö

$$\ln(4 - x) = \ln(x + 8) + 1.$$

log2: Ratkaise eksponenttiyhtälö

$$5^x = 9.$$

log3: Ratkaise logaritmiyhtälö

$$\ln(x) = 10.$$

log4: Ratkaise epäyhtälö

$$\ln(x) < 12.$$

Anna vastauksesi muodossa $x < a$ tai $x > a$, missä a on juuren tarkka arvo.

log5: Ratkaise epäyhtälö

$$3^x > 10.$$

Anna vastauksesi muodossa $x < a$ tai $x > a$, missä a on juuren tarkka arvo.

log6: Mistä muuttujan x arvosta alkaen lauseke

$$\log_6(x)$$

on positiivinen? Anna vastauksesi muodossa $x < a$ tai $x > a$, missä a on juuren tarkka arvo.

log8: Ratkaise yhtälö

$$\log_6(x-2) + \log_6(x-4) = 3.$$

Anna vastaus tarkkana arvona.

C.8 Lukuarvot

n1: Määritä luvun 14 käänteisluku. Anna vastaus tarkkana arvona.

n2: Määritä luvun 28 vastaluku. Anna vastaus tarkkana arvona.

n3: Määritä luvun -8 itseisarvo. Anna vastaus tarkkana arvona.

n4: Laske lausekkeen

$$\frac{d^2}{36} + 3$$

arvo, kun d:llä on arvo 4. Anna vastaus tarkkana arvona.

n5: Laske lausekkeen

$$x^4 + x^3$$

arvo, kun $x = 4$.

n7: Laske lausekkeen

$$-3 - \frac{7}{8} + \left(-\frac{8}{7}\right)$$

arvo. Ilmoita luku murtolukuna X/Y, missä X ja Y ovat kokonaislukuja.

n8 Laske lausekkeen

$$\frac{7}{4} : \frac{-5}{2}$$

arvo. Ilmoita luku murtolukuna X/Y, missä X ja Y ovat kokonaislukuja.

C.9 Potenssit

pow1: Laske seuraavien lausekkeiden arvot

$$(-a)^b, \quad -a^b \quad a^{-b}, \quad -a^{-b},$$

kun $a = 3$ ja $b = 5$.

pow2: Yksinkertaista lausekkeet

$$x^n x^{n+1} (x^{2n})^n \quad \text{ja} \quad \frac{x^n}{x^{n+1}},$$

missä $n = 11$.

pow3: Laske lausekkeiden

$$\sqrt{a}\sqrt{b}, \quad \sqrt{a} + \sqrt{b} \quad \text{ja} \quad \frac{\sqrt{a}}{\sqrt{b}}$$

arvot, kun $a = 81$ ja $b = 64$. Anna vastaukset tarkkoina arvoina.

pow4: Sievennä lauseke polynomiksi

$$(ax^2 - b) + (cx - d)^n (x^m + u),$$

kun $a = 3$, $b = 5$, $c = 3$, $d = -5$, $u = 4$, $m = 3$ ja $n = 2$.

pow5: Määritä lausekkeen

$$\sqrt{4x + 9}$$

nollakohdat. Anna vastaus tarkkana arvona.

pow6: Määritä lausekkeen

$$\sqrt{9x + 27} - 12x + 6$$

reaaliset nollakohdat. Anna vastaus tarkkana arvona.

pow7: Ratkaise yhtälö

$$(x^2 - 3x - 1)^2 = (x^2 - 2x - 4)^2.$$

Vastaukseksi tulee kolme x:n arvoa (joista jotkut voivat olla yhtä suuria).

C.10 Trigonometria

tri1: Kuinka monta radiaania on

$$\alpha = 298$$

astetta? Anna vastaus tarkkana arvona.

tri2: Kuinka monta astetta on 11 radiaania? Anna vastauksesi tarkkana arvona asteina (ilman asteen merkkiä).

tri3: Laske yhtälön

$$\sin(x) = \frac{1}{2}$$

jokin ratkaisu. Anna vastaus radiaaneissa.

tri4: Laske yhtälön

$$\cos(6x) = -1$$

jokin ratkaisu. Anna vastaus radiaaneissa.

tri5: Laske yhtälön

$$\sin(7x) = \cos(x)$$

jokin ratkaisu. Anna vastaus radiaaneissa.

tri6: Mikä on kulman x suuruus, kun

$$\sin(x) - 1 = 0?$$

Anna eräs vastaus radiaaneissa.

tri7: Mikä on kulman x suuruus, kun

$$\tan(2x) = \sqrt{3}?$$

Anna eräs vastaus radiaaneissa.

C.11 Lokitietojen jakautuminen tehtävittäin

Taulukko C.1: Avausten, ratkaisuyritysten, mallivastausten sekä kokonaisklikkausten keskiarvo tehtävittäin.

Tehtävä	Avaus	Yritys	Malliv.	Yhteensä
der2	2,11	1,75	0,27	4,13
der4	1,69	2,05	0,68	4,42
der5	1,64	2,54	1,05	5,23
der6	4,89	3,96	3,91	12,75
der7	1,70	2,64	1,16	5,50
der8	3,54	3,57	2,92	10,03
Jatkuu seuraavalla sivulla				

Tehtävä	Avaus	Yritys	Malliv.	Yhteensä
der9	1,27	1,81	0,40	3,48
der10	1,58	1,55	0,44	3,57
der11	3,60	4,59	3,14	11,34
eq1	1,26	1,37	0,08	2,72
eq2	1,42	2,01	0,60	4,03
eq3	1,42	1,32	0,34	3,08
eq4	1,54	1,63	0,67	3,84
eq5	1,12	1,12	0,19	2,42
eq6	1,54	1,50	0,31	3,36
eq7	1,80	1,92	0,86	4,58
exp1	1,49	2,03	0,22	3,74
exp3	1,23	1,46	0,26	2,95
exp4	1,28	1,66	0,45	3,39
exp5	1,39	1,65	0,64	3,69
exp6	1,25	1,75	0,89	3,90
exp7	1,83	2,25	0,98	5,07
ieq2	1,19	1,65	0,12	2,96
ieq3	1,47	1,64	0,62	3,73
ieq4	1,16	1,55	0,42	3,13
ieq5	1,57	1,76	0,69	4,02
ieq6	4,21	4,22	3,35	11,78
ieq7	4,89	4,04	2,42	11,35
ieq8	1,60	2,61	0,91	5,12
int1	1,56	1,38	0,41	3,35
int2	1,34	1,61	0,46	3,41
int3	1,71	1,85	0,82	4,38
int4	2,64	1,83	2,45	6,92
int5	1,67	2,52	1,42	5,61
int6	4,53	2,99	1,76	9,29
int7	4,65	4,13	4,08	12,85
int8	1,25	1,93	0,54	3,72
lim1	1,22	1,20	0,11	2,53
lim2	1,19	1,42	0,31	2,92
lim3	1,37	1,42	0,53	3,33
lim4	1,43	1,73	1,17	4,34
lim5	1,25	1,51	0,35	3,11
lim6	1,34	1,96	0,89	4,19

Jatkuu seuraavalla sivulla

Tehtävä	Avaus	Yritys	Malliv.	Yhteensä
log1	1,68	1,78	1,61	5,07
log2	1,31	1,61	0,71	3,63
log3	1,18	1,22	0,67	3,07
log4	1,23	1,42	0,43	3,08
log5	1,28	1,73	0,53	3,54
log6	1,24	1,91	1,04	4,19
log8	2,05	1,70	1,64	5,39
n1	1,23	1,34	0,18	2,75
n2	1,10	1,17	0,12	2,39
n3	1,11	1,12	0,14	2,36
n4	1,36	1,99	0,47	3,82
n5	1,36	1,32	0,04	2,73
n7	1,67	2,48	0,36	4,51
n8	1,23	2,09	0,31	3,64
pow1	1,59	1,81	0,26	3,67
pow2	1,53	1,75	0,31	3,58
pow3	1,36	1,56	0,11	3,03
pow4	2,59	3,00	1,47	7,07
pow5	1,20	1,24	0,32	2,76
pow6	1,75	1,75	1,16	4,67
pow7	2,07	2,50	1,41	5,97
tri1	1,31	1,41	0,62	3,34
tri2	1,25	1,75	0,80	3,80
tri3	1,31	1,78	0,96	4,04
tri4	1,81	2,19	1,75	5,76
tri5	1,40	1,38	1,28	4,06
tri6	1,17	1,62	0,72	3,51
tri7	1,24	1,29	0,79	3,31

D. KOLMOGOROVIN–SMIRNOVIN SEKÄ LILLIEFORSSIN TESTIN KRIITTISET ARVOT

D.1 Kolmogorovin–Smirnovin testi

Taulukko D.1: Kolmogorovin–Smirnovin testin testisuureen D kriittiset arvot otoskoolla N ja valitulla merkitsevyystasolla [12].

Otoskoko	Merkitsevyystaso				
N	0,20	0,15	0,10	0,05	0,01
1	0,900	0,925	0,950	0,975	0,995
2	0,684	0,726	0,776	0,842	0,929
3	0,565	0,597	0,642	0,708	0,828
4	0,494	0,525	0,564	0,624	0,733
5	0,446	0,474	0,510	0,565	0,669
6	0,410	0,436	0,470	0,521	0,618
7	0,381	0,405	0,438	0,486	0,577
8	0,358	0,381	0,411	0,457	0,543
9	0,339	0,360	0,388	0,432	0,514
10	0,322	0,342	0,368	0,410	0,490
11	0,307	0,326	0,352	0,391	0,468
12	0,295	0,313	0,338	0,375	0,450
13	0,284	0,302	0,325	0,361	0,433
14	0,274	0,292	0,314	0,349	0,418
15	0,266	0,283	0,304	0,338	0,404
16	0,258	0,274	0,295	0,328	0,392
17	0,250	0,266	0,286	0,318	0,381
18	0,244	0,259	0,278	0,309	0,371
19	0,237	0,252	0,272	0,301	0,363
20	0,231	0,246	0,264	0,294	0,356
25	0,21	0,22	0,24	0,27	0,32
30	0,19	0,20	0,22	0,24	0,29
35	0,18	0,19	0,21	0,23	0,27
yli 35	$\frac{1,07}{\sqrt{N}}$	$\frac{1,14}{\sqrt{N}}$	$\frac{1,22}{\sqrt{N}}$	$\frac{1,36}{\sqrt{N}}$	$\frac{1,63}{\sqrt{N}}$

D.2 Lillieforsin testi

Taulukko D.2: Lillieforsin testin testisuureen D kriittiset arvot otoskoolla N ja valitulla merkitsevyystasolla [11].

Otoskoko	Merkitsevyystaso				
N	0,20	0,15	0,10	0,05	0,01
4	0,300	0,319	0,352	0,381	0,417
5	0,285	0,299	0,315	0,337	0,405
6	0,265	0,277	0,294	0,319	0,364
7	0,247	0,258	0,276	0,300	0,348
8	0,233	0,244	0,261	0,285	0,331
9	0,223	0,233	0,249	0,271	0,311
10	0,215	0,224	0,239	0,258	0,294
11	0,206	0,217	0,230	0,249	0,284
12	0,199	0,212	0,223	0,242	0,275
13	0,190	0,202	0,214	0,234	0,268
14	0,183	0,194	0,207	0,227	0,261
15	0,177	0,187	0,201	0,220	0,257
16	0,173	0,182	0,195	0,213	0,250
17	0,169	0,177	0,189	0,206	0,245
18	0,166	0,173	0,184	0,200	0,239
19	0,163	0,169	0,179	0,195	0,235
20	0,160	0,166	0,174	0,190	0,231
25	0,149	0,153	0,165	0,180	0,203
30	0,131	0,136	0,144	0,161	0,187
yli 30	$\frac{0,736}{\sqrt{N}}$	$\frac{0,768}{\sqrt{N}}$	$\frac{0,805}{\sqrt{N}}$	$\frac{0,886}{\sqrt{N}}$	$\frac{1,031}{\sqrt{N}}$

E. SPEARMANIN KORRELAATIOMATRIISI

Taulukko E.1: Spearmanin järjestyskorrelaatiokertoimet suhteellisella asteikolla mitattavien muuttujien välillä.

Muuttuja	PTT	JäT	Par	IMa	Kli	Ava	Yri	Mv	Oik	Vää
Perustaitotesti	1,00									
Jälkitesti	0,17	1,00								
Parannus	−0,30***	0,86***	1,00							
IMa1u tentti	0,11	0,28**	0,26**	1,00						
Klikkaukset	−0,12	−0,20*	−0,17	−0,23*	1,00					
Avauskerrat	0,03	−0,03	−0,05	−0,08	0,77***	1,00				
Yrityskerrat	−0,15	−0,22*	−0,17	−0,27**	0,70***	0,33***	1,00			
Mallivastaukset	−0,20*	−0,25**	−0,16	−0,28**	0,81***	0,47***	0,52***	1,00		
Oikea vastaus	−0,03	−0,01	−0,00	−0,10	0,29**	0,22*	0,25**	0,33***	1,00	
Väärä vastaus	−0,11	−0,22*	−0,18*	−0,24**	0,64***	0,26**	0,97***	0,45***	0,09	1,00